Open Research Online

The Open University's repository of research publications
and other research outputs

Biological and Clinical Significance of Circulating Tumor Cells in Breast Cancer

Thesis

How to cite:

Fina, Emanuela (2017). Biological and Clinical Significance of Circulating Tumor Cells in Breast Cancer. PhD thesis The Open University.

For guidance on citations see FAQs.

© 2016 The Author

https://creativecommons.org/licenses/by-nc-nd/4.0/

Version: Version of Record

Link(s) to article on publisher's website:
http://dx.doi.org/doi:10.21954/ou.ro.0000c141

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data policy on reuse of materials please consult the policies page.

oro.open.ac.uk

Fondazione IRCCS
Istituto Nazionale dei Tumori
via Venezian, 1 20133 Milano

The Open University (Milton Keynes, UK) - Faculty of Science

PhD Research Degree Programme in Life and Biomolecular Sciences

Affiliated Research Centre:

Fondazione IRCCS Istituto Nazionale dei Tumori – Milan

BIOLOGICAL AND CLINICAL SIGNIFICANCE OF CIRCULATING TUMOR CELLS IN BREAST CANCER

PhD candidate: EMANUELA FINA (Personal Identifier C4842278)

Director of Studies: Dr MARIA GRAZIA DAIDONE

Internal Supervisor: Dr VERA CAPPELLETTI

External Supervisor: Dr ROBERT B. CLARKE

Examiners: Dr CHRISTOPH A. KLEIN

Dr ANGELA GRECO

Chair: Dr LUCA ROZ

Academic Session 2012-2016

Date of submission: October 31, 2016

ABSTRACT

Circulating tumor cells (CTCs) represent a unique source of information that might help clarifying numerous aspects of metastasis biology and finding new clinically relevant biomarkers. On the hypothesis that CTCs possess a distinct profile compared to solid primary and secondary lesions, the transcriptome of experimentally-derived CTCs was compared with those of the primary tumor (PT) and metastases at lymph-nodes (LNs) and lung, in order to identify CTC-specific signatures involved in hematogenous dissemination and which might represent possible prognostic biomarkers.

PTs, CTCs, LNs and lungs were collected from the breast cancer (BC) MDA-MB-231 xenograft model and characterized in two independent experiments using a microarray platform. CTCs were distinguishable from solid lesions by a set of 474 significantly differentially expressed genes. Among genes up-regulated in CTCs, the trefoil factor 3 secreted peptide (*TFF3*) was selected to evaluate its role in CTC biology. *TFF3* down-modulation or knock-out (KO) significantly impaired MDA-MB-231 cell migratory and invasive properties, but not their proliferation rate or vascular-mimicry ability, in *in vitro* assays. Xenograft experiments with MDA-MB-231 *TFF3*^{KO} clones did not allow drawing conclusions on the involvement of TFF3 in dissemination and metastasis as the interpretation of results was hampered by the biological heterogeneity observed among clones. Interestingly, the expression status of *TFF3* and some other CTC/metastasis-specific genes assessed in CTCs isolated from peripheral blood of BC patients, but not CTC status alone defined using standard markers, allowed identifying a group at higher risk of relapse or progression. Indeed, patients with *TFF3*+ CTCs had a significantly shorter progression-free survival compared to those with *TFF3*- CTCs. On the contrary, *TFF3* expression level assessed in publicly available primary BC datasets did not correlate with tumor relapse.

The study of biologically relevant CTC-specific genes may allow deciphering the molecular mechanisms which orchestrate dissemination and real-time monitoring tumor evolution.

ABBREVIATIONS

ADH	atypical ductal hyperplasia
ALH	atypical lobular hyperplasia
AT	AdnaTest
BC	breast cancer
BCCL	breast cancer cell line
CGH	comparative genomic hybridization
CI	confidence interval
CIN	chromosome instability
CSC	cancer stem cell
CTC	circulating tumor cell
CTM	circulating tumor microemboli
DCIS	ductal carcinoma *in situ*
DE	differentially expressed
DTC	disseminated tumor cells
EMT	epithelial-to-mesenchymal transition
EMT1	AdnaTest EMT-1/Stem CellSelect kit
EMT2	AdnaTest EMT-2/Stem CellSelect kit
FACS	fluorescence activated cell sorting
FEA	flat epithelial atypia
FISH	fluorescence *in situ* hybridization
GEP	gene expression profile
GP	22-gene panel
HR	hazard ratio
IDC	invasive ductal carcinoma
IHC	immunohistochemistry
INT	Fondazione IRCCS Istituto Nazionale dei Tumori
IS	isolation support
LCIS	lobular carcinoma *in situ*
LN	lymph-node
LOH	loss of heterozygosity
m.f.p.	mammary fat pad
M+	metastatic

M0	non-metastatic
MET	mesenchymal-to-epithelial transition
n.s.	not significant
OS	overall survival
PBMNC	peripheral blood mononuclear cells
pCR	pathologic complete response
PFS	progression-free survival
PR	partial response
PT	primary tumor
RSN	robust spline normalization
RT	room temperature
SC	ScreenCell
SD	stable disease
T0	time point 0: before starting therapy
T1	time point 1: during therapy
TDLUs	terminal duct and lobular units
TFF	trefoil factor
TN	triple negative
UDH	usual epithelial ductal hyperplasia
UHR	universal human reference
UMR	universal murine reference

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1. BREAST CANCER

1.1.1. The mammary gland development

The mammary gland is structurally a modified and highly specialized sweat gland, which distinguishes mammals from all other animals, and whose function is the production and secretion of milk during lactation. Differently from other glandular organs, the female mammary gland exhibits great plasticity during the lifetime and it reaches full development only during postnatal life. Its development can be subdivided into three major stages: embryonic, pubertal and reproductive (Figure 1.1.1). At present, our knowledge of normal mammary morphogenesis and molecular mechanisms which regulate its development and functions derives primarily from studies performed in mice (Moore et al., 2013; Macias and Hinck, 2012; Cowin and Wysolmerski, 2010). Interaction between the ectodermal epithelium and the mesenchymal cells of the underlying mesoderm is fundamental for tissue architecture, remodeling and functions, and each step of this process undergoes a tiny regulation by selective expression of transcription factors, integration and coordination of signals by endocrine and paracrine molecules and activation of differentiation and proliferation pathways, featuring morphogenetic variation during puberty, pregnancy, lactation, and regression.

During female puberty, the ductal system is stimulated to elongate and further ramify into primary and secondary ducts. At this stage, epithelial ductal cells invade the stroma, a process called branching morphogenesis (Sakakura, 1987) and stratify at their leading edges to form the terminal ductal lobular units - terminal end buds (TEBs) in mice (Cardiff and Wellings, 1999). Ducts consist of an inner layer of luminal epithelial cells surrounded by a layer of myoepithelial cells, which is discontinuous in the small ducts (Figure 1.1.2). With rounds of menstrual cycle the pre-existing ducts continue to branch, as observed by Stute and colleagues in macaques (Stute et al., 2004), making the mammary gland more and more complex.

During pregnancy, the mammary gland undergoes massive tissue remodeling (reviewed in Oakes et al., 2006). The first transformation consists of a tremendous increase in secondary and tertiary ductal branches, providing ductal buds for the second phase, the alveolar morphogenesis (Figure 1.1.2). Rapid and global proliferation of epithelial cells occurs within the ductal branches and developing alveoli (hollow cavities), in order to enlarge the epithelial surface area for milk production (Richert et al., 2000). From mid-pregnancy the gland moves to the secretory phase, where developing alveoli cleave and alveolar cells form a sphere-like milk-secreting single layer (Richert et al., 2000).

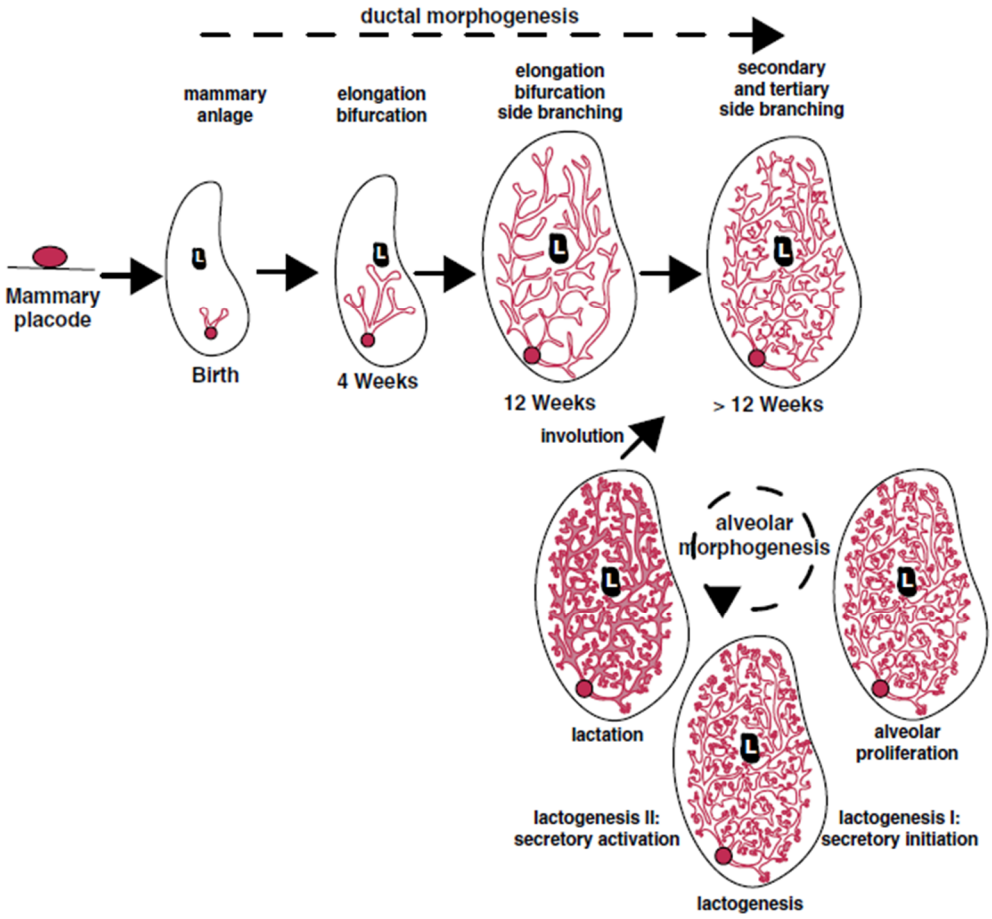


Figure 1.1.1. Schematic representation of the different stages of mammary gland development. At embryonic stage, the mammary placodes differentiates into mammary buds that penetrate the surrounding mesenchyme, elongating and bifurcating in ducts. At puberty the rudimentary ductal tree branches in secondary and tertiary structures. Pregnancy is characterized by alveolar cell expansion, and maturation to milk-secreting acini/alveoli occurs during lactation. Upon weaning, the mammary gland involves to its original adult state. *Oakes et al., 2014*

Each individual alveolus is surrounded by a discontinuous basket-like architecture of myoepithelial cells, and joins to neighbor alveoli to form groups known as lobules. Each lobule is connected with a lactiferous duct and will mature to milk-secreting acini or alveoli during lactation (Oakes et al., 2006). With the cessation of lactation, programmed cell death and extracellular matrix (ECM) remodeling occur, and the mammary gland returns to the virginal phenotype (Macias and Hinck, 2012).

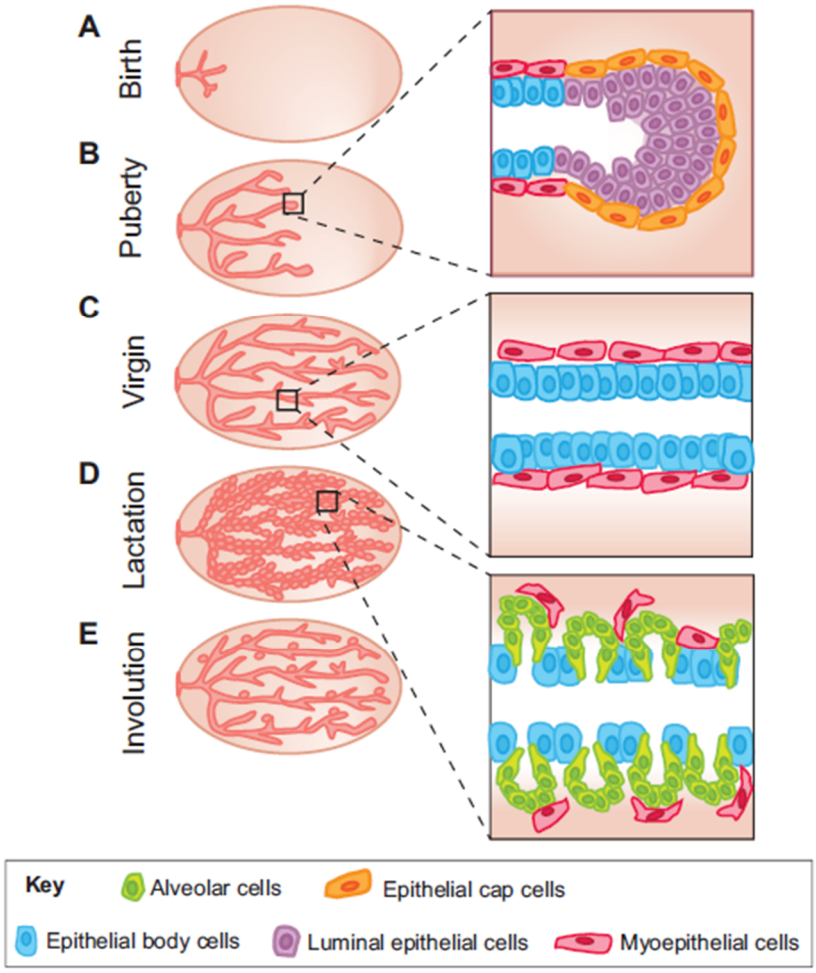


Figure 1.1.2. Cellular bases of mammary gland development. (**A**) The mammary anlage is present at birth and remains quiescent until puberty. (**B**) Epithelial ductal cells expand into the mammary fat pad, led by the terminal end buds, comprising an outer layer of epithelial cap cells surrounded by multi-layered epithelial body cells in rodents. (**C**) The ducts of the adult virgin are characterized by an outer layer of myoepithelial cells and an inner layer of luminal epithelial cells. (**D**) Pregnancy is accompanied by alveolar cell expansion and alveoli morphogenesis. (**E**) After lactation cell death and extracellular matrix remodel the gland to a simple ductal architecture. *Inman et al., 2015*

The adult mammary gland is composed of multiple cell types, including epithelial, adipose, immune, lymphatic and vascular cells and fibroblasts, which interact to maintain the architecture and function of the organ also during specific stages of its development (reviewed in Inman et al., 2015). Since many molecular pathways observed in breast cancer (BC) mimic those regulating mammary cell fate and remodeling in normal breast, studies on mammary gland development are of interest also to cancer biologists and, as epithelial cells are those showing the most dramatic changes during pregnancy and lactation and mammary tumors predominantly arise from the epithelial compartment, recent research efforts have focused on elucidating the hierarchy of epithelial cell differentiation.

1.1.2. Models of mammary cell hierarchy

The observation that the mammary gland is a dynamic organ suggested that a renewable stem or progenitor cell population exists. As such, elucidating normal epithelial differentiation hierarchy is helpful to understand BC heterogeneity and identify the potential cell of origin.

The remarkable regenerative ability of specific cells to reconstitute a mammary gland was first demonstrated by Deome and colleagues (Deome et al., 1959) by transplanting small epithelial fragments of normal or hyperplastic mouse mammary epithelium into de-epithelialized mammary fat pads (m.f.p.). Years later, it was shown that the entire functional mammary gland can be derived from a single cell (Kordon and Smith, 1998; Shackleton et al., 2006), thus supporting the hypothesis that the mammary epithelium contains a stem cell population. Since Deome's demonstration, many studies have focused on identifying and isolating the mammary stem cell (MaSC) population and defining the differentiation potential of different mammary epithelial cell types (reviewed in Oakes et al., 2014, and in Inman et al., 2015).

For half a century, transplantation assay has represented the most common method to identify cells with the ability to reconstitute the glandular epithelium. Significant progress

occurred when mammary tissue dissociation was coupled to fluorescent activated cell sorting (FACS) methods, by which populations with a specific cell surface immunophenotype were injected in the cleared m.f.p. of mice at limiting dilutions to evaluate their ability to form mammary gland-reconstituting units. Moreover, DNA labeling assay was used to identify slow cycling adult stem cells, although the retention of DNA labels might be explained also by asymmetrical segregation of DNA strands (Smith, 2005). FACS analysis allowed the identification and prospective isolation of adult MaSCs based on their $CD24^{+}CD29^{hi}/CD49f^{hi}$ cell surface marker profile (Shackleton et al. 2006; Stingl et al., 2006). Years later, retention of DNA labeling in addition to the expression of $Lin^{-}CD29^{high}CD49f^{high}CD24^{+}$ profile allowed identifying a population with 5-fold higher gland-reconstituting capacity compared to non-labelled/non-label-retaining cells (dos Santos et al., 2013). Further molecular analysis of $Lin^{-}CD29^{high}CD49f^{high}CD24^{+}$ DNA-label-retaining cells revealed that the expression of CD1d (glycoprotein typical of antigen presenting cells) or the lipoprotein receptor-related proteins Lrp5 and Lrp6, as also one of its G-protein coupled receptor downstream effectors Lgr5, improved the selection of cells with regenerating properties compared to the parental population (dos Santos et al., 2013).

Lineage tracing techniques, where a single cell is marked in such a way that the mark is transmitted to the cell's progeny, allowed monitoring the number of founder cells, their location and their differentiation status (reviewed in Kretzschmar and Watt, 2012). This approach, with the advent of genetic recombination and the use of multicolor reports, allows not only better defining the regenerative potential of the traced cell, but also dissecting the mammary epithelial cell (MEC) hierarchy. During the murine mammary gland development the invasion of the m.f.p. by primitive branching structure is accompanied by a concomitant expansion in the MaSC population, also called fetal MaSCs (fMaSCs) by Spike and colleagues (Spike et al., 2012). fMaSCs can be enriched selecting for $CD24^{high}CD49f^{high}$ cells and may presumably overlap with the bipotent MaSC population defined as keratins K14+/K18+ observed in lineage tracing studies (Van Keymeulen et al., 2011; Rios et al., 2014). Indeed, such single-color lineage tracing studies concluded that mammary

development during puberty predominantly occurs through unipotent mammary epithelial progenitors, a luminal and a basal progenitor founder cell (Van Keymeulen et al., 2011; van Amerongen et al., 2012). By contrast, conflicting results emerged from another study by Rios and colleagues (Rios et al., 2014), who demonstrated that the pubertal gland contains several different stem and progenitor cell populations, including bipotent MaSCs, able to generate both the luminal and myoepithelial cells of the duct. Nevertheless, in addition to this bipotent stem cell population, Elf5+ luminal cells labeled before puberty gave rise to luminal and alveolar cells alone, as also observed for K14+/K18+ luminal cells in Van Keymeulen's study.

In a post-pubertal mammary gland, lineage tracing of Elf5+ luminal cells revealed the presence of patches of luminal cells derived from the same progenitor and gradually regressing from week 8 to 20, but when using lineage tracing of K5+ basal cells, both luminal and myoepithelial cells were observed, including some doublets, for long time period (Rios et al., 2014).

Lineage tracing experiments in pregnant mice using K5+, K14+ or Lgr5+ suggested that alveolar cells derive from bipotent progenitor cells passing through the luminal progenitor state first (Rios et al., 2014). Nevertheless, studies based on lineage tracing of Axin2, which is expressed exclusively in basal cells, suggested that alveolar cells arise from a bipotent MaSC instead of a luminal progenitor (van Amerongen et al., 2012). The fate of alveolar cells during mammary gland involution remains unclear. Elf5-labeled luminal progenitors die and are replaced with a new pool of progenitors (Rios et al., 2014), whereas K8- or K18-labeled luminal progenitors persist through multiple cycles of pregnancy (Van Keymeulen et al., 2011).

The molecular mechanisms behind the differentiation of mammary progenitors and/or their progeny into luminal, myoepithelial and alveolar cells are not well understood, but recent transcriptome profiling studies coupled with mouse knock-out models have provided new insights. Crucial driver of luminal progenitor cell differentiation is the transcription factor Gata-3 (Asselin-Labat et al., 2007), whereas potential mediators of the basal phenotype are

p63 (Yalcin-Ozuysal et al., 2010) and MRTF-A (Sun et al., 2006) transcription factors. A cross-talk between the two compartments mediated by Nrg1 (neuregulin 1), which is expressed in myoepithelial cells in a p63-dependent manner, was observed (Forster et al., 2014). Nrg1 binds luminal cells via Erbb4 receptor and induces the expression of the Stat5a targets Elf5 and cyclin D1, necessary for luminal and luminal progenitor cell function (Forster et al., 2014).

The mammary gland is thought to house multiple populations of stem cells that fulfill the requirement of self-renewal and differentiation into mature cell lineages. At present, according to the most widely accepted hypothetical model of the mammary epithelial hierarchy (Visvader and Stingl, 2014; Figure 1.1.3), a multipotent fetal MaSC gives rise to a heterogeneous stem cell compartment, comprising long-term and short-term repopulating cells, from which luminal and basal precursors originate.

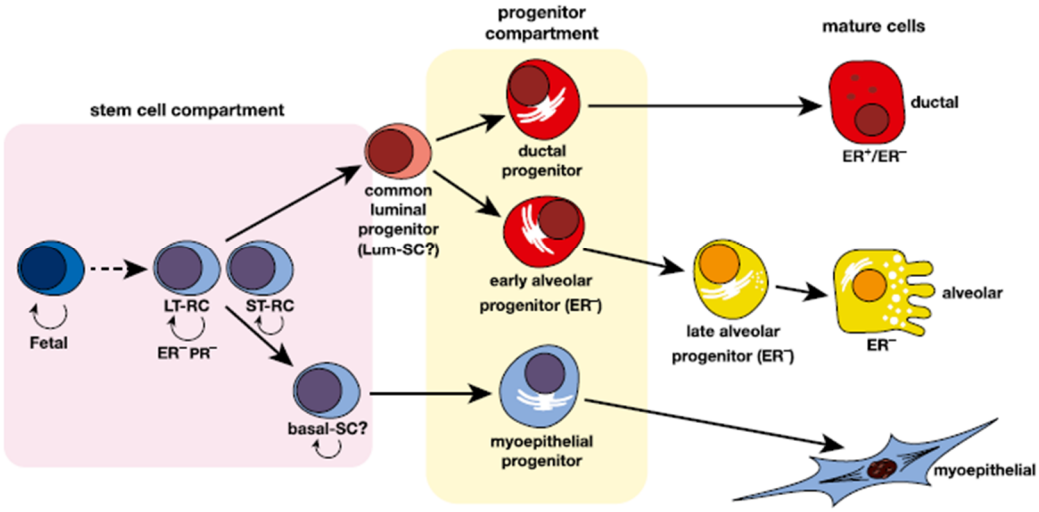


Figure 1.1.3. Hypothetical model of the mammary epithelial hierarchy. A multipotent fetal MaSC has been identified. In the adult mammary gland, the stem cell compartment is heterogeneous and appears to comprise long-term and short-term repopulating cells (LT-RCs and ST-RCs, respectively), both of which are multipotent. These in turn give rise to committed progenitor cells for the myoepithelial and luminal (ductal and alveolar) epithelial lineages, but the precise number of progenitor cells is yet to be determined. Luminal progenitors are restricted to either a ductal or alveolar cell fate. The ductal progenitor possibly comprises both hormone receptor (HR)-positive and HR-negative cells, while the early and late alveolar-restricted progenitors are likely to be HR-negative. There may be a common luminal progenitor for these sublineages. The prospective isolation of cellular subsets from mouse and human mammary tissue provides support for the depicted hierarchical organization. In addition, two types of unipotent cells (lum-SC and myo-SC) may exist; current lineage tracing data is also consistent with long-lived progenitors performing these functions *in vivo*. *Visvader and Stingl, 2014*

One of the most controversial issues arising from *in situ* lineage tracing studies is related to the identity and the role of stem cells in the adult mammary gland homeostasis (Figure 1.1.4). One model suggests that the luminal and basal compartments are maintained by lineage-restricted, unipotent stem cells (Van Keymeulen et al., 2011; van Amerongen et al., 2012); the other proposes that the adult mammary epithelium is maintained by bipotent stem cells (Rios et al., 2014). To address the issue of lineage identity and developmental plasticity, Granit and colleagues have recently proposed an alternative to the traditional MEC hierarchy. They suggest a multidimensional classification of MECs along several distinct axes, including stem cell versus differentiated, basal versus luminal and mesenchymal versus epithelial identity. Such multidimensional model allows description of intermediate and mixed differentiation states and can be applied to both normal mammary gland and BC. Cancer cells possessing a high degree of stemness would display increased capacity to shift between positions on such a multidimensional scale, and to acquire intermediate phenotypes on its different axes (Granit et al., 2014).

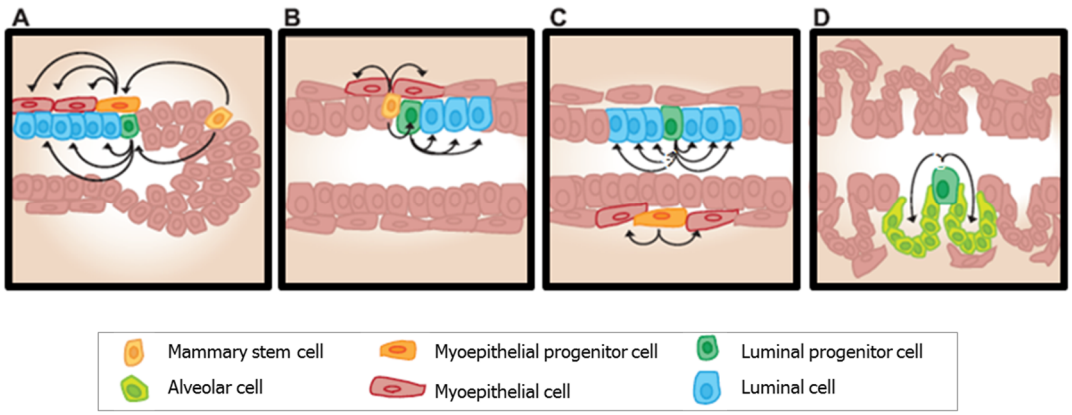


Figure 1.1.4. Contrasting models of mammary stem cell hierarchy across breast development stages. (**A**) During puberty mammary stem cells are present in the terminal end bud. They are able to self-renew and differentiate into mature cell lineages. (**B**) Other evidence from the adult virgin mouse gland suggests that bipotent stem cells can give rise both to luminal and myoepithelial cells. (**C**) By contrast, other lines of evidence suggest that unipotent progenitor cells contribute to ductal maintenance in the adult gland, thus myoepithelial progenitors are thought to differentiate directly into myoepithelial cells. (**D**) Finally, during pregnancy, alveolar progenitor cells in the luminal compartment give rise to the secretory alveolar cells. *Inman et al., 2015*

1.1.3. Epidemiology of breast cancer

Breast cancer (BC) is the most frequently diagnosed cancer and the leading cause of cancer death in females worldwide, accounting for 23% (1.38 million) of the total new cancer cases and 14% (458,400) of the total cancer deaths, according to GLOBOCAN 2008 estimates (Ferlay et al., 2010; Jemal et al., 2011). About 50% of BC cases and 60% of related deaths are estimated to occur in economically developing countries, as a result of population aging and growth, as well as an adoption of cancer-associated lifestyle choices including smoking, physical activity and food habits (Jemal et al., 2011). Incidence rates are high in Western and Northern Europe, Australia/New Zeland and North America, while intermediate and low incidence can be observed in South America, the Caribbean and Northern Africa, and in sub-Saharan Africa and Asia, respectively (Jemal et al., 2011). Such international variation in incidence rates is largely due to differences in reproductive and hormonal factors (Jemal et al., 2010), as in alcohol consumption (Key et al., 2006). During the 1990s both incidence rates strikingly increased in many Western countries, in concomitance with the introduction of screening by mammography (Jemal et al., 2011). In contrast, mortality has been declined in North America and several European countries over the past 25 years, largely as a result of early detection and improved treatment (Jemal et al., 2011).

At present, many factors increasing the risk of developing disease have been clarified, while others are still under investigation, and efforts to identify strategies for preventing BC have been made.

One of the most documented risk factors for BC is age (Singletary, 2003). The incidence is extremely low before age 30, after which increases linearly until the age of 80. As also reported for many types of cancers, a strong link exists between aging and the accumulation of damages at molecular level, as a result of less effective cellular repair mechanisms, which may explain why cancer is typically a disease of advanced age (DePinho, 2000).

The association between BC risk and reproductive factors has been widely investigated, and it still represents a matter of study since it strongly supports a hormonal role also in BC etiology (reviewed in Key et al., 2001). Early menarche and late menopause increase risk of developing BC. Early age at first term birth is related to lifetime incidence reduction and women who had at least one full-term pregnancy benefit from around 25% reduction, compared with nulliparous women. Childbearing seems to have a dual effect as risk increases in the period immediately after birth, but it gradually diminishes with time and, in the longer term, it is protective. Furthermore, protection increases with the number of full-term pregnancies. In addition to nulliparity, the absence or short-lifetime duration of breastfeeding, which is typical of developed countries, substantially contributes to the high incidence of BC. Also studies in less developed countries, where the total duration of breastfeeding can be very long, have reported a protective effect. Such observations might be explained by the occurrence of cell differentiation during lactation and subsequent lower susceptibility of epithelial stem cells to carcinogenesis.

Use of exogenous hormones was also linked to the risk of developing BC as a meta-analysis of 51 epidemiologic studies by the Collaborative Group on Hormonal Factors in Breast Cancer (1997) demonstrated a 26% increase in the risk of BC over a 5-year period. Importantly, the risk reduces after use of hormonal replacement therapy is stopped and disappears after about 5 years. Researchers have also been looking to the use of oral contraceptives for many years, but results of these studies are still controversial.

At the present time there is growing interest in the possible health threat posed by endocrine-disrupting chemicals (EDCs), which are substances in our environment, food, and consumer products that interfere with hormone biosynthesis, metabolism, or action, resulting in a deviation from normal homeostatic control or reproduction. Indeed, some of them are chemicals with structural similarities to estrogens, and there is evidence that endocrine disruptors have effects on male and female reproduction, breast development and cancer (reviewed in Diamanti-Kandarakis et al., 2009).

Importantly, family history of the disease represents a key risk factor, indicating that genetic determinants also contribute to the incidence of BC. Indeed, 10% to 30% of BC cases can be attributed to hereditary factors (reviewed in Apostolou and Fostira, 2013). The two most important BC susceptibility genes are *BRCA1* and *BRCA2* tumor suppressors, identified by linkage analysis and positional cloning in the 1990s as key regulators of DNA repair, transcription, and cell cycle in response to DNA damage (Miki et al., 1994; Wooster et al., 1995; Ralhan et al., 2007). Germline mutations in the *BRCA1* gene have been associated with the triple-negative (TN) BC subtype, accounting for 60% to 80% of BCs from *BRCA1* mutation carriers (Atchley et al., 2008).

Over the past several years, our knowledge of specific genetic defects that may contribute to familial BC has significantly increased thanks to genome-wide-association studies, and beside *BRCA1/2*, other genes showing variable penetrance are now associated with BC genetic risk, *e.g. TP53*, *PTEN*, *STK11/LKB1*, *CDH1*, *CHECK2*, *PALB2*, *ATM*, *BRIP1*, *RAD51C*, *XRCC2*, *NBS1*, *RAD50*, *MRE11*, *BARD1*, *ABRAXAS*, *RAD51D*, *MAP3K1*, *FGFR2*, *LSP1*, *TNRC19* and *H19* (reviewed in Mavaddat et al., 2010; Apostolou and Fostira, 2013).

Recent advances in DNA sequencing have led to the development of BC susceptibility gene panels for germline genetic testing. Risk assessment models and criteria for cancer genetics referral as well as guidelines for screening, surveillance protocols and prophylactic options were reported by the American Cancer Society and amended by the National Comprehensive Cancer Network (reviewed in Stuckey et al., 2016; www.ncnn.org). A major limitation of genetic testing is represented by the number of inconclusive results due to variants of unknown significance, as missense and splice-site mutations or silent variants or mutations in the exonic splice enhancers, thus further efforts should be made to better understand their contribution to the pathogenesis of BC. The potentials of next-generation sequencing (NGS) technologies may accelerate the process towards the discovery of new susceptibility genes and their defects, providing the route to more precise genetic counseling and new targeted therapies.

1.1.4. On the molecular origin of breast cancer

Several lines of evidence indicate that tumorigenesis is a multistep process in which all tumor cells derive from a single ancestral cell that acquired genetic and epigenetic changes providing strong clone-specific selective advantages (Nowell, 1976; Greaves and Maley, 2012). Such a process is the result of many factors, including inherited mutations or polymorphism of cancer susceptibility genes, environmental agents that influence the acquisition of somatic genetic changes, and several other systemic and local factors, such as hormones and inflammation (Ponder et al., 2001; Polyak, 2001; Coussens and Werb, 2002).

According to the so-called model of Darwinian somatic evolution of cancer (Nowell, 1976; Fearon and Vogelstein, 1990), it was assumed that cancer proceeds as a linear succession of clonal expansions of cells with progressively increased fitness, triggered by the acquisition of advantageous driver mutations. In line with this model, colorectal cancer has for more than two decades served as the paradigm for the multi-step concept of cancer initiation and progression, theorized by Fearon and Vogelstein (1990).

A major challenge to BC researchers has been and continues to be the ability to distinguish genetic alterations that are critical to tumor initiation from those that are epiphenomena of genetic instability, which is a hallmark of tumor progression (Hanahan and Weinberg, 2011). Sporadic BC, which constitutes more than 90% of all BCs, is a complex and heterogeneous disease at both the clinical and molecular levels. Understanding this heterogeneity is crucial for the development of effective therapies and adequate management of patients.

The genomic diversity present in cancer cells ranges from single nucleotide changes to large-scale cytogenetic alterations and is caused by increased genomic instability (Lengauer et al., 1998). Two main classes of genomic instability exist, each one giving rise to nucleotide mutations or complex chromosomal rearrangements (chromosomal instability, CIN). There are many mechanisms at the basis of CIN, including multipolar spindles, improper chromosome condensation or cohesion, defects in mitotic spindle assembly/dynamics, defective mitotic checkpoint and telomere attrition, replication stress, and improper

kinetochore-microtubules attachments (Gordon et al., 2012; Thompson et al., 2010). Telomere dysfunction drives chromosome fusion and breakage-fusion-bridge (BFB) cycles, resulting in complex, unbalanced chromosome rearrangements (Artandi and DePinho, 2000), a phenomenon also called “telomere crisis” observed by *in situ* analysis during the transition from mammary ductal hyperplasia to ductal carcinoma *in situ* (DCIS) (Chin et al., 2004).

One of the main products of CIN is aneuploidy, a condition associated with the gain or loss of whole chromosomes or parts, thereof, leading to genomic imbalances (Geigl et al., 2008; Gordon et al., 2012). In BC, DNA aneuploidy is detectable in up to a third of patients with atypical hyperplasia of mammary epithelium (Carpenter et al., 1987), and it is present in the majority of patients with invasive BC carcinoma (Aasmundstad and Haugen, 1990). CIN has been suggested to provide phenotypic variation and increase tumor heterogeneity, therefore fuelling the ability of cancer cells to progress and adapt to chemotherapy (Roschke and Rozenblum, 2013; McGranahan et al., 2012). However, the relationship between CIN and drug resistance is far from simple. Indeed, by stratifying tumors using a CIN expression signature, Swanton and colleagues found that BCs with the lowest or the highest CIN signatures were associated with improved prognosis compared to those with intermediate scores (Birkbak et al., 2011).

Studies based on fluorescence *in situ* hybridization (FISH) and conventional comparative genomic hybridization (CGH) analysis (Gray et al., 1994; Courjal and Theillet, 1997) describe recurrent DNA amplifications (with presumptive oncogene driver) at 8p12 (*FGFR1*), 8q24 (*MYC*), 11q13 (*CCND1*), and 17q12 (*ERBB2*). Another type of genetic abnormality frequently observed in breast tumors is the loss of heterozygosity (LOH) on chromosomes and chromosome arms 1, 3p, 6q, 7q, 8p, 11, 13q, 16q, 17, 18q, and 22q (Bièche and Lidereau, 1995).

Along with the development of array-based CGH protocols and whole-genome sequencing (WGS), several studies defined the fine structure of BC genomes, providing a high resolution catalogue of recurrent alterations in correlation with the molecular classification based on transcriptome analysis (Cancer Genome Atlas Network, 2012;

Nik-Zainal et al., 2016). Beside to genes previously described as implicated in BC (*PIK3CA*, *PTEN*, *AKT1*, *TP53*, *GATA3*, *CDH1*, *RB1*, *MLL3*, *MAP3K1*, *ERBB2*, *MYC* and *CDKN1B*), such studies allowed the identification of a number of novel mutated genes (*TBX3*, *RUNX1*, *CBFB*, *AFF2*, *PIK3R1*, *PTPN22*, *PTPRD*, *NF1*, *SF3B1*, *CCND3*, *MED23*, *FOXP1*, *MLLT4*, *XBP1* and *ZFP36L1*), harboring somatic mutations supposed to be implicated in BC oncogenesis, defined as driver mutations (Stratton et al., 2009).

A clear picture of the molecular pathogenesis of BC has not been drawn yet, due to the intrinsic heterogeneity of this neoplasm. Since increasing genetic instability is a hallmark of tumor progression, Kornelia Polyak sustains and even provided evidence that understanding at the molecular level the pathophysiology of lesions in early stage cancers, such as DCIS, could be useful to identify critical events in the development of BC and could lead to the development of novel therapies that decrease the incidence of potentially lethal invasive cancers (Polyak, 2001).

1.1.5. Genes and signaling pathways involved in breast cancer initiation and progression

Cancer development is a gradual and complex process resulting from the disruption of molecular pathways involved in a series of biological processes (Hanahan and Weinberg, 2011).

In breast cancer, the functional implication of the human epidermal growth factor receptor 2 (HER2/*neu* or ERBB2) aberrant expression was first demonstrated two decades ago (Slamon et al., 1987) and numerous data now support its transforming potential. The rodent *neu* oncogene, so called because discovered in a rodent neuroblastoma cell line, was demonstrated to be the same ortholog of the human *ERBB2* (Coussens et al., 1985). The transforming function in the *neu* oncogene is conferred by a point mutation within the transmembrane domain, resulting in a V664E mutated protein named neuT that promotes receptor dimerization and enhances tyrosine kinase activity (Bargmann et al., 1986; Weiner

et al., 1989). Numerous mouse transgenic models have also confirmed the role of this oncogene in tumorigenesis since the activated *neu* oncogene (neuT) expressed in mice mammary tissue (MMTV-neuT mice) induces adenocarcinomas (Muller et al., 1988; Bouchard et al., 1989).

The relevance of the experimental data to human disease is also supported by a substantial body of clinical data. Overexpression of the HER2 protein, either through gene amplification or through transcriptional deregulation, is seen in approximately 25-30% of breast and ovarian cancers, and it confers worse biological behavior (Slamon et al., 1989; Burstein, 2005). Breast cancers can have up to 25–50 copies of the *HER2* gene, and up to 40–100-fold increase in HER2 protein resulting in 2 million receptors expressed at the tumor cell surface (Kallioniemi et al., 1992). Nevertheless, recent genome-sequencing works allowed identification of *HER2* somatic mutations in BCs lacking *HER2* gene amplification, some of which were demonstrated to be activating mutations, suggesting that they likely drive tumorigenesis (Bose et al., 2013).

ERBB2 proto-oncogene encodes for a 185 kDa transmembrane glycoprotein belonging to the family of four human epidermal growth factor receptors, designated HER1 to HER4, respectively. A ligand for HER2 has not been identified and it is hypothesized that the main function of HER2 is to co-operate in signal transduction by forming heterodimers with other members of the HER family (Graus-Porta et al., 1997; Figure 1.1.5). Several mechanisms of HER2-mediated tumorigenesis have been proposed (reviewed in Moasser, 2007), and in the most widely accepted mechanistic model HER2 induces transformation through increased kinase activity (reviewed in Harari and Yarden, 2000).

Numerous studies have shown an intriguing interrelationship between HER2 and estrogen receptor (ER), and to a lesser extent, progesterone receptor (PgR) expression in BC. With exceptions, there is a strong, highly significant inverse relationship between ER status and either HER2 or HER1 overexpression (Battaglia et al., 1988; Cappelletti et al., 1988) and numerous studies also indicate that a mutually repressive feedback signaling loop exists between HER2 and ER expression (reviewed in Harari and Yarden, 2000).

Nevertheless, even estrogen, working via the non-genomic activity of the membrane-associated ER, has been shown to activate HER2 signaling (Shou et al., 2004).

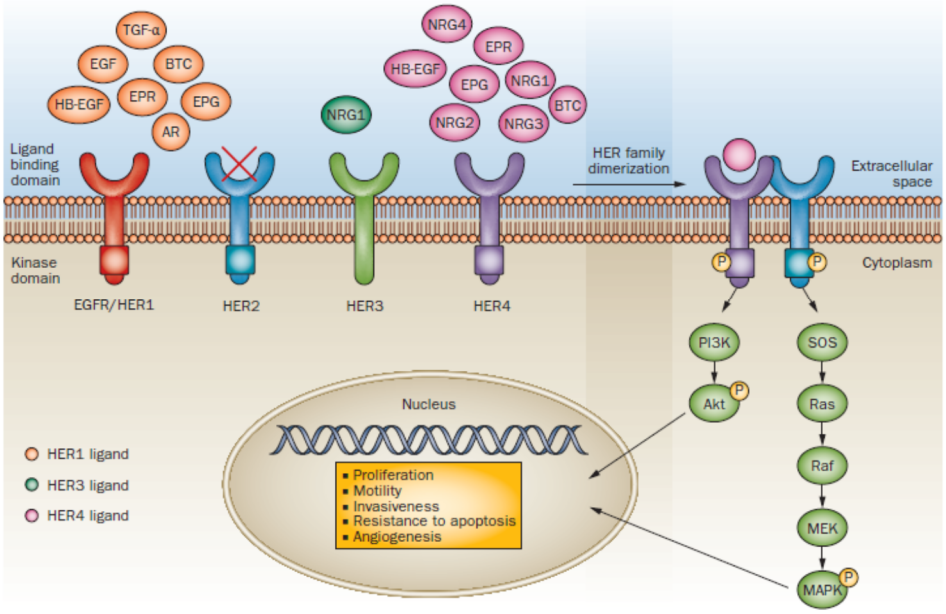


Figure 1.1.5. ERBB2 heterodimerization and signaling pathway. Signaling downstream of HER family activation is dependent on heterodimerization of the HER family member triggered by ligand binding to the extracellular ligand-binding domain (with the exception of HER2, which has no identified ligand and is always in an open conformation that allows dimerization). Phosphorylation of the HER kinase domains (with the exception of HER3, which does not have a kinase domain) initiates a downstream cascade resulting in *VEGF* transcription and other physiological responses required for carcinogenesis. Abbreviations: AR, amphiregulin; BTC, betacellulin; EPG, epigen; EPR, epiregulin; HB-EFG, heparin-binding EGF-like ligand; NRG, neuregulin. *Arteaga et al., 2011*

The connection between a woman's hormonal status and her risk of developing BC has been suspected since 1713, when Bernardino Ramazzini noted that nulliparous were more likely to develop BC than women who had borne children. In 1896 George Beatson provided evidence for the first time for the dependence of BC on ovarian function. He observed the regression of both advanced cancer (Beatson, 1896) and metastatic disease (Boyd, 1900) in premenopausal women following oophorectomy.

Estrogen effects are exerted through two types of specific nuclear receptor, estrogen receptor alpha (ERα, *ESR1*, chromosome 6) and beta (ERβ, *ESR2*, chromosome 14), which act in a ligand-dependent manner. The binding of estrogen in the hormone-binding domain (HBD), induces a trans-conformational change of the whole molecule allowing unmasking of

the activating function 1 (AF-1) in domain A/B, with subsequent dimerization, translocation to the nucleus and binding to estrogen-responsive element (ERE, reviewed in Sommer and Fuqua, 2001).

Beside the classical model, estradiol (E_2) modulates transcription via a non-canonical activation pathway, according to which ligand binding leads to the interaction of ER with other transcription factors as AP-1, Sp1 or NF-κB. A number of genes, including those encoding for trefoil factor 1/pS2, cathepsin D, cyclin D1, gene regulated by estrogen in breast cancer 1 (*GREB1*), c-Myc and progesterone receptor, are positively regulated by ERα, as also confirmed for the majority of them by chromatin immunoprecipitation studies (reviewed in Welboren et al., 2007).

Another mechanism, originally called “non-genomic” (Morley et al., 1992), has been implicated in the ER activity, since a rapid activation of several signaling pathways (PLC/PKC, Ras/Raf/MAPK, PI3K/Akt and cAMP/PKA) was observed to occur upon estradiol binding. The plasma membrane-associated ER is responsible for non-genomic pathway and co-operates with ER’s nuclear activity (Björnström and Sjöberg, 2005; Figure 1.1.6).

Regulation of ER activity is also mediated in a ligand-independent manner (reviewed in Marino et al., 2006) by a series of signaling pathways downstream of growth receptors, such as EGFR, IGFR and ERBB2, which induce activation of the receptor via phosphorylation of some serine (for example S118, Kato et al., 1995) and tyrosine residues in the AF-1 and AF-2 domains. In addition, mutations in the ERα gene *ESR1*, such as the first described Tyr 537 Asn (Zhang et al., 1997) at the HBD level, and others more recently identified (Robinson et al., 2013), have been shown to elicit changes in receptor activation by conferring constitutive activity.

Ample clinical and experimental data demonstrates the major role played by ER in BC development and progression, so that pharmacological targeting of the receptor has become an important treatment strategy, also for chemoprevention. Antiestrogens (AEs) are used successfully in order to inhibit ER-mediated activation of gene transcription although,

unfortunately, clinical resistance develops eventually (reviewed in Clarke et al., 2003, and in Viedma-Rodríguez et al., 2014).

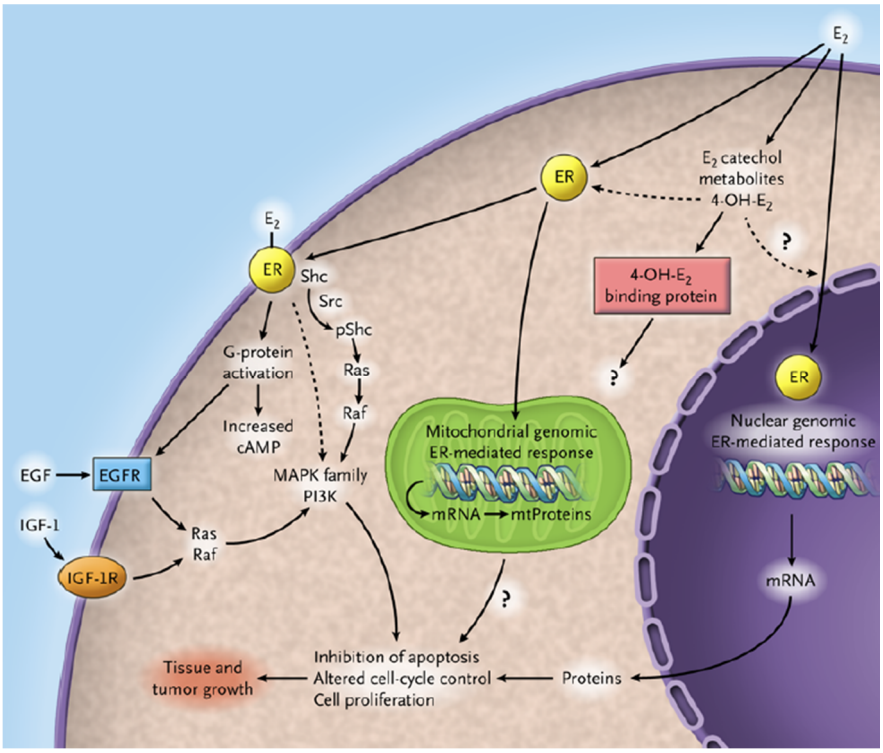


Figure 1.1.6. Estrogen receptor signaling pathway. ER response to estradiol and its metabolites is exerted by transcription of genes prevalently related to cell proliferation, and by a non-genomic pathway which involves a cytosolic or membrane-located form of ER. Its activity is also mediated in a ligand-independent manner, *via* the activation of tyrosine kinase receptor pathways. Abbreviations: ER, estrogen receptor; EGFR, epidermal growth factor receptor; IGF-1R, insulin growth factor receptor 1; E_2, 17-β-estradiol. *Yager and Davidson, 2006*

During the last decade much attention has been focused on targeting the receptor tyrosine-kinases (RTKs) signaling pathway, which is aberrantly activated in BC with overwhelming frequency (Alvarez et al., 2010). RTKs, such as EGFR, IGFR, FGFR and VEGR, can activate downstream Ras/Raf/MAPK, JAK/Stat, PI3K/AKT/mTOR, JNK, and PLCγ signalling pathways, which regulate a plethora of cellular functions such as growth, survival, proliferation, metabolism, apoptosis, adhesion and motility (Figure 1.1.7).

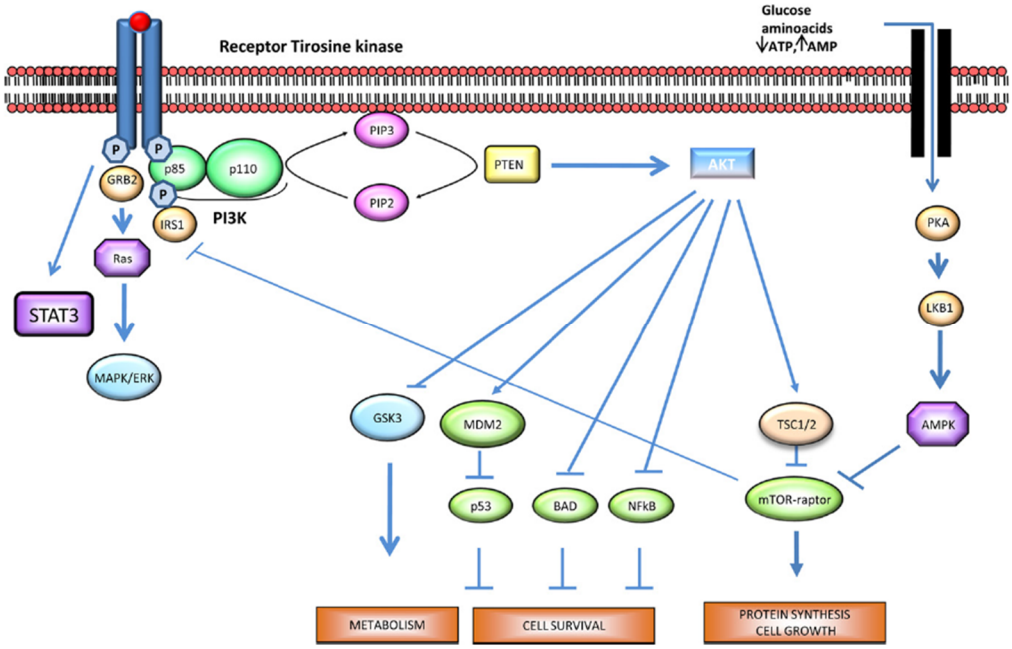


Figure 1.1.7. Deregulated signaling pathways in breast cancer. *Eroles et al., 2012*

IGFR has been reported to be involved in BC progression. Fifteen percent of BCs contain genomic alterations in the IGF pathway, documented by NGS (Cancer Genome Atlas Network, 2012), which consist mainly of amplification and are generally rare, with only IGF1R and IRS2 showing amplification in >5% of cases, whereas when considering mRNA levels 45.3% of BCs show a molecular alteration in at least one IGF family member. IGF pathway is regulated in many critical points, from ligand availability to negative feedback mechanisms exerted by mTOR (Wan et al., 2007), and IGF1R/PI3K/AKT/mTOR pathway has been largely looked into the past decade as a target to treat BC (Tabernero et al., 2008).

Amplification of FGFR genes, including *FGFR1* and *FGFR2*, was initially documented in human BC samples in the early 1990s (Adnane et al., 1991). A surge of studies within the last 5-10 years has both confirmed these initial observations and expanded significantly upon the mechanisms through which the FGF/FGFR axis contributes to BC (Hynes and Dey, 2010).

Apoptosis plays a key part in the development of the normal breast and its deregulation overrides many of the normal checkpoint pathways and leads to the expansion of neoplastic

cells. Recent data suggest that a subset of triple-negative/basal-like breast cancer cells is sensitive to TRAIL as a single agent (Rahman et al., 2009). In addition, many studies have demonstrated that resistance to TRAIL-mediated apoptosis in BC cells can be overcome by combinations of TRAIL with chemotherapy, radiation, and various targeted agents. Moreover, cyclin D1 has been reported to be overexpressed in more than 50% of human BC. Its expression causes mammary cancer in transgenic mice, altered *CCND1* expression contributes to the loss of normal cell cycle control during tumorigenesis and recent studies have demonstrated that cyclin D1 conducts additional specific functions to regulate gene expression at chromatin level, promotes cellular migration and inhibits mitochondrial metabolism (reviewed in Velasco-Velázquez et al., 2011).

One feature of *BRCA*-mutated cancers is the defective function of one of the major DNA damage repair pathways, the homologous recombination (HR) pathway. The observation that *BRCA*-mutated BCs show impairment in HR pathways and that some sporadic triple-negative BCs (TNBCs) display a phenotype resembling *BRCA1*-mutated cancers without harboring a *BRCA1* mutation (Turner et al., 2004) led to the application of PARP inhibition, as enzyme involved in DNA-damage repair, for TNBC treatment (reviewed in Livraghi and Garber, 2015).

Finally, recent discovery and revaluation of the portion of human genome non-coding for proteins opened a new window in cancer studies. More interestingly, it has been noticed that also alterations of non-coding genes, including micro-RNAs and long non-coding RNAs (lncRNAs), are related to BC pathogenesis. Altered micro-RNA expression has been observed in BC, and a large body of evidence proved a contribute for some of these gene expression modulators, such as miR-21, miR-155, miR-10b, miR-34a, miR-125a,b, and the miR-200 family, also to the development of cancerous phenotypes (reviewed in van Schooneveld et al., 2015). Accumulating evidence also highlights the potential role of lncRNAs as biomarkers and therapeutic targets in solid tumors. Recently, compelling results in preclinical transgenic models proved that targeting *MALAT1* could represent a safe and effective treatment for BC (Arun et al., 2016). Moreover, the first profiling study on 658 cases

of infiltrating ductal carcinomas of the breast from The Cancer Genome Atlas project allowed identification of lncRNAs that are likely to play important role as regulators of BC initiation and progression, probably affecting gene expression by chromatin remodeling and histone modification (Su X. et al., 2014).

1.1.6. Histological pathogenesis of breast cancer

The current model of human BC progression proposes a linear multistep process which initiates as flat epithelial atypia (FEA) - characterized by replacement of the native epithelial cells of the terminal duct and lobular units (TDLUs) by one and up to five cell layers of mildly atypical cuboidal to columnar cell population with apical snouts - progresses to atypical ductal hyperplasia (ADH) – when the ducts are only partially or largely populated by the atypical cells, *i.e.* cells with morphological and differentiation alterations -, evolves into low to high grade ductal carcinoma *in situ* (DCIS) - presence of mildly to highly atypical cells with low to high-grade, pleomorphic nuclei with irregular contour -, and culminates in the potentially lethal stage of invasive ductal carcinoma (IDC), in a sort of evolutionary continuum (Oyama et al., 1999; Page and Rogers, 1992).

The ductal and lobular histological subtypes constitute the majority of all BCs worldwide, with the ductal subtype accounting for 50-80% of all diagnosed cases (Ellis, 2003; Lakhani et al., 2012). Epidemiologic and morphologic observations led to the formulation of several linear models of BC initiation, transformation and progression (reviewed in Bombonati and Sgroi, 2011; Figure 1.1.8). For the ductal subtype, two models have been proposed. The first “ductal” model recognizes FEA, ADH and DCIS as the non-obligate precursors of invasive and metastatic ductal carcinoma. The second “ductal” model, supported by epidemiological studies, proposed usual epithelial ductal hyperplasia (UDH) - characterized by proliferation of a heterogeneous cell population with irregularly shaped and sized secondary lumens - as an intermediate stage of progression between FEA and DCIS. Immunohistochemical and recent molecular biological evidence strongly suggests that UDH

is not a precursor to ADH and that this second model of progression is likely to be invalid. For the lobular subtype, the progression scheme recognizes atypical lobular hyperplasia (ALH) and lobular carcinoma *in situ* (LCIS) as the non-obligate precursor lesions to invasive lobular carcinomas (reviewed in Bombonati and Sgroi, 2011).

Additional support for these models is provided by several genome and transcriptome analyses. CGH studies demonstrated frequent loss of 16q in low-grade DCIS and 13q loss and amplification of 17q12 and 11q13 in high-grade DCIS. LOH-based studies identified loss of 16q as a hot spot in ADH and a recurrent chromosomal aberration in FEA, and revealed that this alteration is most frequently shared with low-grade rather than high-grade DCIS. Notably, this common pattern of genomic alterations is not observed in UDH, lesion that displays rare and randomly distributed chromosomal alterations that make it more similar to the histopathologically normal breast than ADH. Until recently, identification of the precursor lesion of high grade DCIS remained elusive (reviewed in Bombonati and Sgroi, 2011).

Together, these studies provide evidence that: 1) DCIS, like IDC, consists of two distinct genetic pathways correlating with tumor grade, and that DCIS is direct precursor to IDC, 2) ADH is a precursor to low-grade DCIS, 3) FEA is genetically related to ADH and is likely a precursor to ADH, 4) at the genomic level UDH is most similar to normal breast epithelium and is not a precursor to ADH. In addition, immunohistochemistry (IHC) and comparative genomic hybridization (CGH) studies suggest that 5) microglandular adenosis may represent a precursor lesion to a subset of high grade DCIS lesions (reviewed in Bombonati and Sgroi, 2011).

Comparative gene expression studies support the concept that low-grade and high-grade DCIS likely arise from two distinct evolutionary pathways and that ADH is the precursor of low-grade DCIS, since ADH and low-grade DCIS share a near identical gene expression profile enriched in genes associated with the ER-positive phenotype, whereas high-grade DCIS possesses uniquely different profile enriched in genes associated with mitotic-activity and cell-cycle progression (Ma et al., 2003). In addition, the study demonstrated that globally no consistent major transcriptional changes occurred between the

preinvasive and invasive stages and that the transition from preinvasive disease to invasive disease was mainly associated with quantitative, rather than qualitative, differences. This indicates that BC progression may be more intricate than predicted by the traditional model based on activation of oncogenes and inactivation of tumor suppressor genes, and that this process may be reliant upon quantitative and temporal gene expression shifts.

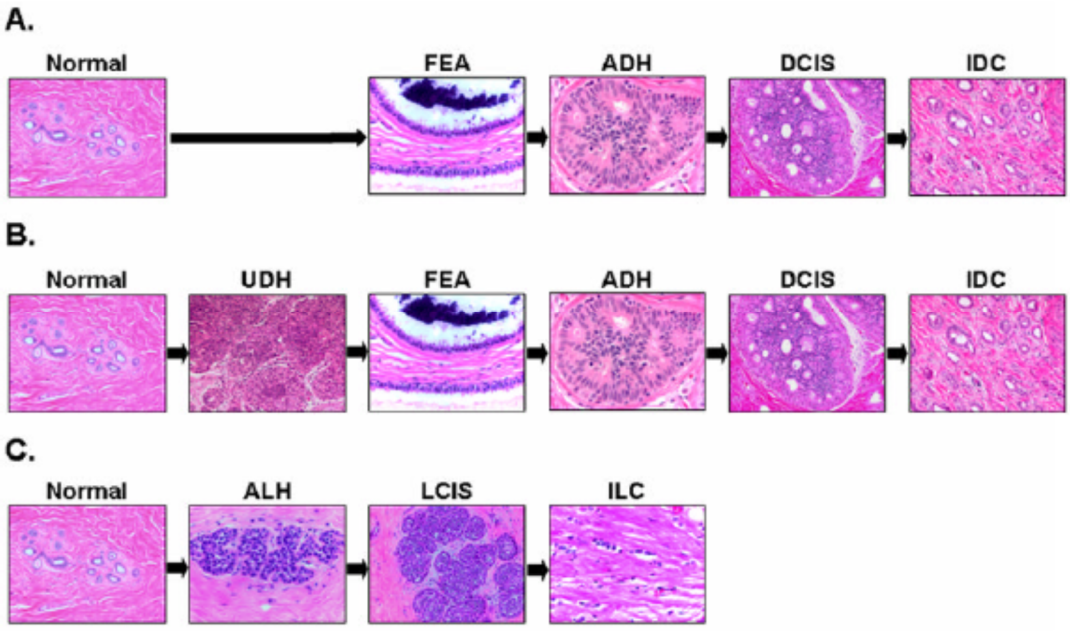


Figure 1.1.8. Models of breast cancer progression. (**A**) One of the "ductal" models recognizes FEA, ADH and DCIS as the non-obligate precursors of invasive and metastatic ductal carcinoma. (**B**) Another "ductal" model proposes usual epithelial ductal hyperplasia (UDH) as an intermediate stage of progression between FEA and DCIS. (**C**) Progression model for the lobular subtype, according to which atypical lobular hyperplasia (ALH) and lobular carcinoma *in situ* (LCIS) are the non-obligate precursor lesions to invasive lobular carcinoma. *Bombonati and Sgroi, 2011*

Differently from DCIS, few studies have focused on the genomic alterations associated with LCIS. These studies demonstrated common loss of chromosome 16q in ALH, LCIS and ILC supporting an evolutionary link among the three lesions (Mastracci et al., 2006; Morandi et al., 2006).

Recent advances in comprehensive, high-throughput genome, transcriptome and epigenome analyses has provided a more complete understanding of the complex genetic

and biological inter-relationships of the different stages of human BC evolution. Human BC appears to progress along two distinct molecular genetic pathways that strongly associate with tumor grade. Moreover, the cancer stem cell (CSC) model and the role of the tumor microenvironment as not simple bystander of the cancerogenesis process contribute to the complexity of BC. Despite these significant advances, we have only begun to understand the pathogenesis of BC and it is anticipated that our knowledge in this field will receive novel insights with the advent of additional novel technologies.

1.1.7. Breast cancer heterogeneity and biomarkers

The classification of invasive BC currently involves the assessment of histological criteria encompassing both morphology-based and IHC analyses. Traditional pathological parameters such as histological type, tumor size, histological grade and axillary lymph-node (LN) involvement have been shown to correlate with clinical outcome and provide the basis for prognostic evaluation (Elston et al., 1999). IHC markers such as the expression of hormone receptors ER and PgR and the overexpression and/or amplification of HER2 provide additional therapeutic predictive value and are of key importance in guiding treatment selection (Harris et al., 2007).

Hormone receptor-positive breast cancers account for around 75–80% of all cases and standardized IHC assays for the routine testing of ER and PgR are used to guide the selection of patients for hormonal-based therapies. HER2 represents the only additional predictive marker currently in routine use. Approximately 10–15% of breast cancers have *HER2* overexpression and/or amplification with around half of these co-expressing hormone receptors (Konecny et al., 2003). These patients are selected for anti-HER2 based therapies, including the humanized monoclonal HER2 antibody, trastuzumab, which targets the extracellular domain of the HER2 receptor. The remaining 10–15% of BCs are defined by hormone receptor and HER2 negativity (*i.e.*, TNBCs), which represent a key clinical entity given their lack of therapeutic options (Dawson et al., 2009).

While the current classification of human breast tumors has been fundamental for prognostic and predictive evaluation, there remain a number of important limitations. First, considerable variation in response to therapy and clinical outcome still exists, even for tumors with apparent similarities in clinical and pathological characteristics. Second, this classification continues to provide limited insight into the complex underlying biology and the molecular pathways driving the disease in different subtypes.

Expression analysis using microarray-based technology has provided researchers with an opportunity to begin moving towards comprehensive molecular profiling of breast cancer. These efforts have resulted in the identification of clinically relevant molecular subtypes, and have provided early insights into the molecular heterogeneity of the disease (Perou et al., 2000; Sorlie et al., 2003; Hu et al., 2006). Five distinct intrinsic subtypes have been identified based solely on gene expression: luminal A, luminal B, HER2 overexpressing, basal-like and normal breast tissue-like. Differences in gene expression patterns reflect basic alterations in the cell biology of the tumors and importantly are associated with significant variation in clinical outcome (Sorlie et al., 2003). The prognosis of patients with ER-positive disease is largely determined by the expression of genes related to proliferation (Hu et al., 2006). More recently, the intrinsic classification has been refined in a PAM50 assay based on the expression of 50 genes designed to classify single samples into each of the five intrinsic subtypes (Parker et al., 2009; Nielsen et al., 2010).

Following the initial identification of the intrinsic molecular subtypes, gene expression studies have evolved and further sub-classification of BCs into new molecular entities have been proposed. For example, a detailed analysis of genes differentially expressed in ER-negative tumors has demonstrated that basal breast cancers are a heterogeneous group with at least four main subtypes (Teschendorff et al., 2007). Furthermore, this analysis revealed an immune response gene expression module, which identifies a good prognosis subtype in ER-negative disease. Other recent studies have also identified a new breast cancer intrinsic subtype known as Claudin-low or mesenchymal-like (Prat et al., 2010). This subtype is characteristically negative for ER, PgR and HER2 and carries an intermediate

prognosis between basal and luminal subtypes. Importantly, Claudin-low/mesenchymal tumors appear to be enriched with cells showing distinct biological properties associated with mammary stem cells and tumor initiating potential (Hennessy et al., 2009; Lim et al, 2009; Lehmann et al., 2011).

Distinct patterns of genomic rearrangements in BC have been characterized and changes in gene expression patterns were shown to be influenced by the genomic architecture (Chin et al., 2006; Ding et al., 2010). The emergence of NGS technologies has now allowed the characterization of the mutational landscape of the disease. These analyses have identified novel cancer genes found to be recurrently mutated in BC and have demonstrated the extent of heterogeneity across BC genomes (Shah et al., 2009; Shah et al., 2012; Nik-Zainal et al., 2016). Through the integrated analysis of both genomic and gene expression data across large numbers of BCs, Curtis and colleagues have recently extensively characterized 2000 breast tumors as part of METABRIC (Molecular Taxonomy of Breast Cancer International Consortium), revealing the existence of 10 novel molecular subgroups (Curtis et al., 2012), called integrative clusters (IntClust 1–10), each associated with distinct copy number somatic aberrations (CNAs) and gene expression changes. These clusters clearly demonstrated the heterogeneity present within tumors classified according to ER, PgR and HER2 expression, and they divided all of the previously identified intrinsic subtypes into separate groups (Figure 1.1.9). Furthermore, the 10 groups were associated with distinct clinical features and outcomes (Curtis et al., 2012).

In parallel with the identification of the intrinsic subtypes, gene expression profiling (GEP) has also been used by several groups to identify distinct prognostic signatures (van de Vijver et al., 2002; van't Veer et al., 2002; Paik et al., 2004). Two of these signatures, MammaPrint, a microarray-based assay of 70-gene breast cancer signature developed in Amsterdam, and OncotypeDX, a polymerase chain reaction (PCR)-based assay of a panel of 21 genes, have been approved for clinical use and are now being tested in randomized clinical trials (Cardoso et al., 2008; Sparano and Paik, 2008).

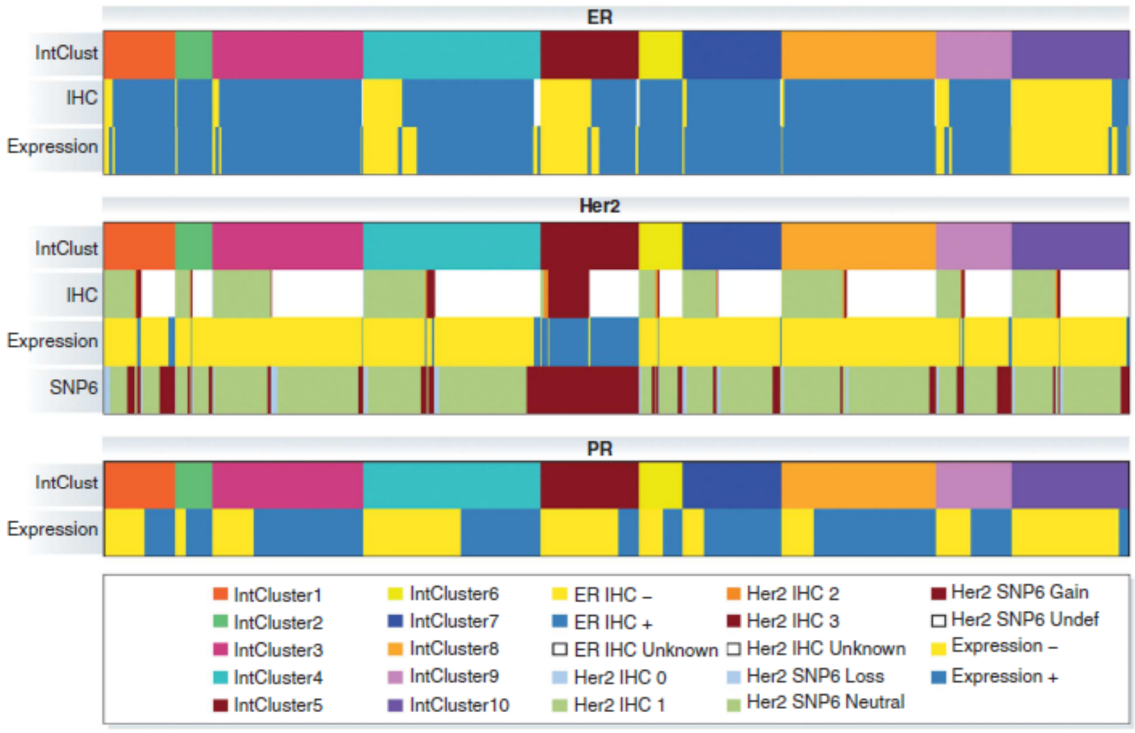


Figure 1.1.9. Relationship between the 10 integrative clusters and ER, HER2 and PgR expression. *Dawson et al., 2013*

MammaPrint is a Food and Drug Administration (FDA)-cleared microarray-based test that uses expression levels of the 70 genes to assess distant recurrence risk in early-stage BC. The primary analysis of the MINDACT trial to prospectively test the 70-gene signature was conducted on patients deemed clinically high risk but genomically low risk who were randomized not to receive chemotherapy. The 5-year distant metastasis–free survival (DMFS) for this group was 94.7%, thus confirming the value of genomic profiling for patients with early BC with 0 to 3 positive LNs (Piccart et al., 2016). Moreover, the prospective RASTER study proved that MammaPrint Low Risk patients can safely forgo chemotherapy (Drukker et al., 2013), which is further subject of the prospective randomized MINDACT trial. The 21-gene expression assay was proven to predict recurrence in tamoxifen-treated node-negative BC patients (Paik et al., 2004) and benefit from adjuvant chemotherapy in estrogen-receptor–positive disease (Paik et al., 2006), and its usefulness was furthermore validated in the prospective TAILORx study (Sparano et al., 2015). However, a recent report

showed that intratumor heterogeneity may affect the detection of genes included in various GEP signatures (Oncotype Dx, MammaPrint, PAM50, EndoPredict, and Breast Cancer Index), and therefore the ability of the test to predict prognostic risk (Gyanchandani et al., 2016).

Monitoring tumor-specific alterations in cell-free DNA (or circulating tumor DNA) is now under investigation as tumor 'liquid biopsy'. Circulating tumor DNA (ctDNA) may be used to characterize tumor heterogeneity and metastasis-specific mutations providing on tumor adaptation during therapy and helping clinical decision-making when a shift toward alternative and personalized therapies is needed. Recent studies provided a proof of concept of the potential utility of ctDNA monitoring, demonstrating that tracking specific mutations in blood allows early measurement of response to treatment in metastatic BC (Dawson et al., 2013) and early prediction of relapse in early BC (Garcia-Murillas et al., 2015).

1.1.8. Breast cancer diagnosis and treatment

(Senkus et al., Primary BC:ESMO clinical practice guidelines, Ann Oncol 2015; www.cancer.gov; AIOM guidelines www.aiom.it)

The diagnosis of BC is based on clinical examination in combination with imaging, and it is confirmed by pathological assessment. Other assessments are complete personal and family medical history. Clinical examination includes bimanual palpation of the breasts and locoregional LNs and assessment for distant metastases (bones, liver and lungs; a neurological examination is only required when symptoms are present). Imaging includes bilateral mammography and ultrasound of the breast and regional LNs.

At present, bilateral mammography is considered the most effective screening test and many European countries have established national or regional population-based mammography screening programmes to detect BCs at a pre-clinical stage. Indeed, screening every 2 years has shown the greatest mortality reduction benefit in the 50- to

70-year age group, while the evidence for effectiveness of mammography screening in women aged 40-49 years is limited. In women with familial BC, with or without proven *BRCA* mutations, annual screening with magnetic resonance imaging (MRI) of the breast, in combination with mammography is recommended, but it should be considered also in cases of breast implants, lobular cancers, suspicion of multifocality/multicentricity (particularly in lobular BC) or large discrepancies between conventional imaging and clinical examination.

Apart from physical examination and imaging, it is mandatory to perform pre-treatment pathological examination of the primary tumor histology and cytology of the axillary nodes, if involvement is suspected. Indeed, BC cells are most likely to spread first to LNs located in the axilla, even if in cancers near the breastbone they tend to spread first to LNs inside the chest before they can be detected in the axilla. Lymph node involvement occurs at different extent. Isolated tumor cells are defined as small clusters of cells not greater than 0.2 mm, or nonconfluent or nearly confluent clusters of cells not exceeding 200 cells in a single histologic LN cross section. Micrometastatic nodal involvement is defined as a cluster of cells >0.2 mm but no greater than 2.0 mm. Macrometastatic involvement of the axillary nodes (classically designated as "node-positive") is defined by any tumor cell cluster >2.0 mm. The presence of macrometastases within the axillary nodes is a well-established and independent prognostic factor, with a worse prognosis associated with greater nodal involvement. Occult micrometastases refer to nodal metastases that are not seen on initial histological examination but are detected subsequently by IHC or PCR.

Pathological diagnosis provides information useful for tumor staging, histological classification, grading and IHC evaluation of ER status (using a standardized assessment methodology, *e.g.* Allred or H-score), PgR status and HER2 expression, as also of Ki67 labeling index. *HER2* gene amplification status may be determined directly on all invasive tumors using *in situ* hybridization (fluorescent, chromogenic or silver), in addition to IHC or only for tumors with an ambiguous (2+) IHC score. The guidelines for HER2 testing have recently been updated by the ASCO–College of American Pathologists group. A change has been introduced in the definition of HER2 positivity by IHC (tumor are defined 3+ when more

than 10% of the cells, instead of 30%, show a complete membrane staining), and by *in situ* hybridization (positive if the number of *HER2* gene copies is ≥6 or the ratio HER2/chromosome 17 is ≥2, instead of 2.2). The definition of equivocal cases is broader; if a case is defined as equivocal after two tests it is eligible for trastuzumab, and should be discussed in multidisciplinary tumor boards (Wolff et al., 2013).

As asymptomatic distant metastases are very rare, assessment of metastatic disease is based on physical examination. Additional investigations such as chest computed tomography (CT), abdominal ultrasound or CT scan and bone scan are considered for patients with clinically positive axillary nodes, large tumors (*e.g.* ≥5 cm), aggressive biology or clinical signs, symptoms or laboratory values suggesting the presence of metastases. Dual imaging methods combining functional and anatomical information such as the ^{18}F-fluorodeoxyglucose (^{18}F-FDG) Positron Emission Tomography (PET)/CT may be useful when conventional methods are inconclusive.

Patients with BC have different treatment options: surgery, radiation therapy, chemotherapy, hormone therapy and targeted therapy. The choice of treatment strategy is based on biology (pathology including biomarkers, gene expression) and tumor extent/location (size and location of primary tumor, number of lesions, number and extent of lymph node involvement), as well as on the age and menopausal status.

Local treatment includes surgery and radiation therapy (RT). The major change in the surgical treatment of primary BC has been a shift from radical mastectomy towards breast conservation, which started more than 30 years ago (Veronesi et al.,1977, Veronesi et al., 1981; Veronesi et al., 2002).

Regional lymph node status remains one of the strongest predictors of long-term prognosis in primary BC. Historically, removal of tumor positive lymph nodes (called axillary lymph node dissection) was done to help disease staging and prevent regional recurrence, *i.e.* cell migration to nearby lymph nodes that give rise to a new tumor. Since removing multiple lymph nodes at the same time has been associated with adverse effects, as lymphedema, sentinel lymph node biopsy (SLNB) procedure has been introduced in order to

identify, thanks to injection of a tracer, and remove only lymph nodes involved in cancer spread (Veronesi et al., 1997).

Whole breast RT is recommended after breast-conserving surgery (BCS) to reduce the risk of recurrence (including locoregional and distant). Radiation after mastectomy is recommended for high-risk patients, including those with involved axillary nodes and/or T_3-T_4 tumors. Locoregional RT encompassing the chest wall and all regional lymph nodes is indicated when cancer has spread in the lymph nodes or at distant sites, such as the bones or brain.

Diverse systemic treatments for BC are available, and their choice is essentially tailored on the phenotypic subtypes, determined by ER/PgR, HER2 and Ki67 assessment, or on the molecularly-based intrinsic subtype.

For tumors with ER-positive and/or PgR-positive cancer cells, whose growth is affected by the level of circulating hormones, endocrine therapy (ET) represents the treatment of choice. It is based on the administration of hormones analogues that modulate the function (tamoxifen, toremifene) or cause degradation (fulvestrant) of their receptors. Aromatase inhibitors (AIs), as letrozole, anastrozole and exemestane, are drugs that interfere with estrogen production in post-menopausal women, where a small amount of estrogen is still made by aromatase enzyme present in the fat tissue. In pre-menopausal women also luteinizing hormone-releasing hormone (LHRH) analogs are used to induce temporary ovarian ablation.

Patients with HER2-positive BC are eligible for targeted therapy with a recombinant humanized monoclonal antibody directed against the extracellular subdomain IV of HER2 (trastuzumab). The effect of trastuzumab on tumor proliferation is mediated by several mechanisms of actions (Hudis, 2007), such as internalization and degradation of HER2, inhibition of MAPK and PI3K-Akt downstream signaling pathways, antigen-dependent cellular cytotoxicity mediated by natural killer cells, and inhibition of tumor angiogenesis. Clinical studies have shown that the combination of trastuzumab with standard chemotherapy produces far better response rates than chemotherapy alone (Slamon et al., 2001; Vogel et

al., 2002). Thus, the combinations that include trastuzumab have been considered as the standard of care for HER2-overexpressing BC patients (Hudis, 2007).

In addition to trastuzumab, other anti-HER2 biological therapies have been approved (reviewed in Baselga and Swain, 2009). Pertuzumab is a monoclonal antibody which binds to HER2 subdomain II and acts by inhibiting HER2 heterodimerization with HER1, HER3 and HER4. A recent clinical trial (CLEOPATRA) showed that first-line therapy with pertuzumab, trastuzumab, and docetaxel significantly improved overall survival among patients with HER2-positive metastatic BC, as compared with placebo, trastuzumab, and docetaxel as control groups (Swain et al., 2015).

Nevertheless, therapy resistance may occur and since women with advanced or metastatic HER2-positive BC have limited therapeutic options once their disease has progressed on trastuzumab-based standard initial chemotherapy regimens, novel anti-HER2 therapies have been explored. A new drug is represented by a small molecule (lapatinib), acting as a dual receptor tyrosine kinase inhibitor of both ErbB1 and ErbB2. In addition, lapatinib combined with capecitabine has demonstrated superior efficacy over capecitabine alone in this group of patients compared to the monotherapy arm (Geyer et al., 2006), and now it is indicated in combination with capecitabine for the treatment of patients with advanced or metastatic BC whose tumors overexpress HER2 and who have received prior therapy including an anthracycline, a taxane, and trastuzumab.

Also surgery and systemic therapies can be combined according to the stage of the disease. In just the past decade, there has been a rapid evolution in BC treatment protocols. With the primary objective to improve surgical outcomes in patients for whom a primary surgical approach is technically not feasible and in patients with operable BC who desire breast conservation, but for whom either a mastectomy is required or a partial mastectomy would result in a poor cosmetic outcome, neoadjuvant, or perioperative, therapy, *i.e.* the systemic treatment of BC prior to definitive surgical therapy, (Kaufmann et al., 2006; Gralow et al., 2008), has been introduced into clinical practice. Neoadjuvant treatment has been compared with standard, postoperative adjuvant chemotherapy, with the dual goals of

improving survival and facilitating local therapies. Unfortunately, neoadjuvant chemotherapy does not seem to improve overall survival, as demonstrated in the National Surgical Adjuvant Breast and Bowel Project B18 trial, among others (reviewed in Schott and Hayes, 2012). Aside from the potential clinical benefits that are achieved by downstaging, neoadjuvant therapy allows direct and early observation of the response to treatment, which in theory could lead to modifications of the treatment plan in the event of poor response. However, clinical and radiographic monitoring during neoadjuvant chemotherapy to predict pathologic complete response (pCR) is notoriously inaccurate and there is little agreement regarding the precise definition of pCR. On the basis of the limited clinical advantages of neoadjuvant chemotherapy, postoperative adjuvant systemic therapy is considered the standard of care (Schott and Hayes, 2012). Nevertheless, excluding these clinical objectives, neoadjuvant therapy gives researchers the opportunity to obtain tumor specimens (both fresh and formalin-fixed) and blood samples prior to and during the preoperative treatment in order to identify tumor- or patient-specific biomarkers (Daidone et al., 2011).

Although most patients present with localized breast cancer and may be cured with local therapy, distant recurrence may occur and is the primary cause of death from the disease. The widespread use of adjuvant therapies, including both systemic therapies and RT given after the primary treatment, proved to be effective in reducing the risk of distant and local recurrence, and has contributed to reduced BC mortality rates (reviewed in Anampa et al., 2015). In 1976, Bonadonna and colleagues published the results of their landmark trial of adjuvant chemotherapy for breast cancer (Bonadonna et al., 1976), showing that 12 months of postoperative chemotherapy consisting of cyclophosphamide, methotrexate, and fluorouracil (CMF) decreased the risk of recurrence of BC in women with positive axillary lymph nodes. Since then, adjuvant cytotoxic chemotherapy regimens have evolved from single alkylating agents to polychemotherapy regimens incorporating anthracyclines and/or taxanes and treatments are recommended to start within 2-6 weeks after surgery. Subsequent trials also showed benefit in lower risk post-menopausal women (Albain et al., 2009) and women with axillary node-negative disease (Mansour et al., 1989; Fisher et al.,

1997). Despite the most recent publication of the Early Breast Cancer Trialists' Collaborative Group overview states that the relative benefit of chemotherapy is similar in all the subgroups independent of age, stage, histopathological grade and ER status (Peto et al., 2012), there is large consensus with the 2013 and 2015 St. Gallen guidelines, which recommend, for luminal cases with unclear chemotherapy indications, that the decision on systemic adjuvant therapies should be based on the surrogate intrinsic phenotype (Figure 1.1.10), that is determined by ER/PgR, HER2 and Ki67 assessment, with the help of genomic tests when available (Coates et al., 2015). Generally, i) all patients with detectable ER expression, defined as ≥1% of invasive cancer cells, should be offered ET, ii) for luminal HER2-negative cancers, chemotherapy should not be used concomitantly with ET, iii) while Luminal B HER2-positive tumors should be treated with chemotherapy, ET and trastuzumab; iv) HER2-positive (non-luminal) cancers should be treated with chemotherapy plus trastuzumab, and v) triple-negative tumors benefit from adjuvant chemotherapy only, consisting of four to eight cycles of anthracycline- and/or taxane-based regimen, and available alternatives to treat these patients are still scarce.

Patients presenting with metastatic disease at first diagnosis represent only the 7% of the population. In the vast majority of cases stage IV BC is diagnosed in patients previously treated with neoadjuvant therapy. About 30% of node-negative and 70% of node-positive patients will eventually undergo disease relapse, within shorter time for patients with TN and HER2-positive BC, and with a higher risk of bone metastases in patients with hormone-sensitive tumors, or recurrence at central nervous system in women with HER2-positive and triple-negative tumors. The aim of treatment for secondary (metastatic) BC is to control the growth or spread of cancer, to relieve symptoms and improve or maintain the quality of life. Also in this case biological features of disease are important in treatment choice, aside with clinical parameters to distinguish an indolent from an aggressive metastatic disease. In the presence of osteolytic lesions, bisphosphonates are used to reduce the risk of fractures and studies are ongoing to evaluate if they improve the benefit of adjuvant chemotherapies. Overall, treatment options for patients with recurrent or metastatic

disease are limited and remain a major therapeutic challenge (Santa-Maria and Gradishar, 2015).

New therapies have been developed in recent years, and some of them were also approved. In 2007, the U.S. FDA approved ixabepilone – which stabilize microtubules - for the treatment of aggressive metastatic or locally advanced breast cancer no longer responding to currently available chemotherapies. Ixabepilone, in combination with capecitabine, has demonstrated effectiveness in the treatment of metastatic or locally advanced breast cancer in patients after failure of an anthracycline and a taxane.

The mesylate salt Eribulin (fully synthetic macrocyclic analogue of halichondrin B which acts on microtubules) was also approved by the FDA in 2010 to treat patients with metastatic BC who have received at least two prior chemotherapy regimens for late-stage disease, including both anthracycline- and taxane-based chemotherapies.

Everolimus (mTOR inhibitor) is approved for post-menopausal women with advanced hormone receptor-positive, HER2-negative BC. It is used along with the aromatase inhibitor exemestane (Aromasin) for women progressing under letrozole or anastrozole treatments.

PARP inhibitors are most likely to be helpful against cancers caused by BRCA mutations. These drugs have shown some promise in early clinical trials treating some types of breast and other cancers. Further studies are being done to determine when these drugs might be most helpful. Mitoxantrone is a type II topoisomerase inhibitor; also used in metastatic BC.

In 2015 an oral inhibitor of CDK4/6 (palbociclib) received FDA approval for the treatment of ER-positive HER2-negative advanced or metastatic BC in combination with fulvestrant in women with disease progression following endocrine therapy (Turner et al., 2015).

Based on the success of immunotherapeutic agents in the treatment of melanoma, and more recently in lung cancer, it is expected that immunotherapeutic strategies will be proven efficacious for the treatment of patients with many other solid tumor types.

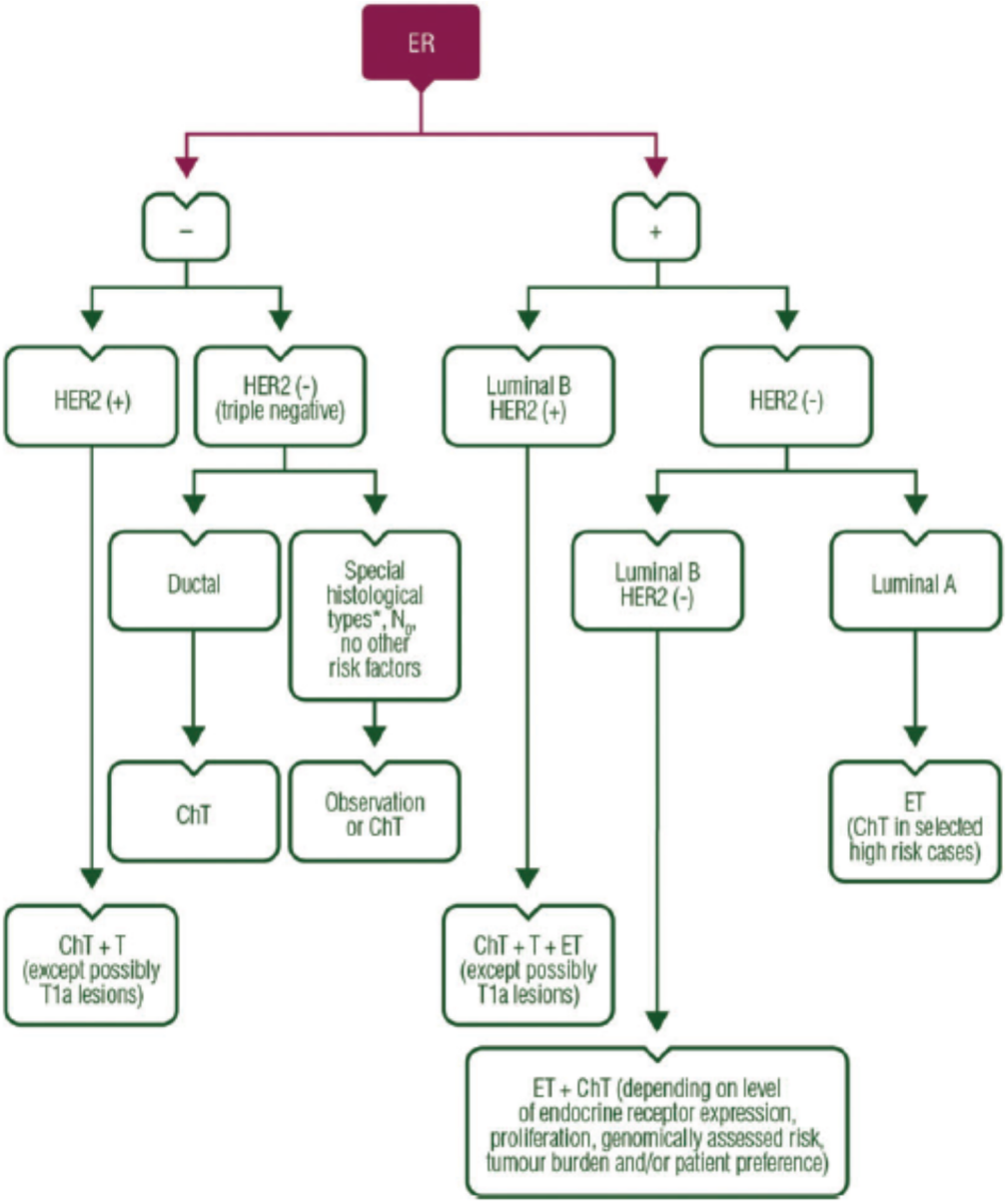


Figure 1.1.10. Schematic treatment strategy in operable breast cancer. (Neo)adjuvant systemic treatment choice is tailored on biomarker expression and intrinsic phenotype. ER, estrogen receptor; HER2, human epidermal growth factor 2 receptor; ChT, chemotherapy; ET, endocrine therapy; T, trastuzumab*. Senkus et al., 2015*

In BC, the relevance of the host immune response to the tumor has long been debated. The prognostic significance of tumor-infiltrating lymphocytes (TILs) in early BC has been reported in numerous studies (reviewed in Ocaña et al., 2015). A robust body of literature now suggests that BC, particularly the more aggressive subtypes of HER2-positive and TNBC, does elicit host antitumor immune responses, and that the robustness of the

response correlates with prognosis (Adams S. et al., 2014; Ali et al., 2014; Loi et al., 2014). Therefore, there is great interest in exploring the potential role of immunotherapy in treating patients with BC. In addition to monoclonal antibodies and some vaccines strategies, large interest has been recently shown for immune checkpoint inhibitors. Data were recently reported from two trials of antibodies (pembrolizumab and atezolizumab) targeting the T-cell inhibitory molecule PD-1 or its ligand, programmed cell death receptor 1 ligand (PD-L1), two negative regulators of immune checkpoint (Muenst et al., 2013; Nanda et al., 2016), showing promising results in terms of safety and treatment response.

1.1.9. Breast cancer-initiating cells

The term “stem cells” was used for the first time by Edmund Beecher Wilson in 1896 (Wilson, 1896) and was associated for about a century to normal embryonic and adult tissue hierarchy and development. Only in most recent years, the use of the term “stemness” has been applied also to cancer biology, although the multilineage differentiation potential is ascribable to normal stem cells only, while cancer stem cells (CSCs) are not known to differentiate into more cell types, except into cells forming the bulk of the tumor. However, in this context, CSCs can be located at the helm of tumor hierarchy and possess traits of self-renewal, *i.e.* the ability to divide asymmetrically yielding daughter cells that inherited the parental cell’s characteristics, thus remaining CSCs, as well as daughter cells that differentiate into the neoplastic cells forming the remaining part of the tumor.

The phenotypes of CSCs are complex and strictly dependent on the type of tumor. For this reason, participants in The Year 2011 Working Conference on CSCs have outlined guidelines on how to provide a conceptual classification of these cells depending on the biological system in which they are being studied (Valent et al., 2012). Operational definition of CSCs comprises three main features: 1) they express a specific set of surface markers that allow their identification and isolation, 2) they are able to grow in non-adherence

conditions, and 3) they can rebuild the tumor of origin when injected in immunocompromised mice even at limiting dilution (Dalerba et al., 2007).

Hypotheses regarding the origin of CSCs include 1) malignant transformation of pre-existing normal stem cells, 2) mature cancer cell dedifferentiation with EMT and 3) induced pluripotent cancer cells (reviewed in Islam et al., 2015). This implies that rather than a stable entity at the top of a hierarchical model, CSCs are in a dynamic status as they can continually dedifferentiate from progenitors or differentiated cancer cells (Wicha et al., 2006; Visvader and Lindeman, 2008; Ni and Huang, 2013).

The presence of stem-like cells in human BC was first demonstrated in 2003 (Al-Hajj et al., 2003). They can be propagated in serum-free and non-adherence culture conditions as mammospheres and exhibit a $CD44^{+}CD24^{low/-}$ immunophenotype (Al-Hajj et al., 2003; Dontu et al., 2003; Ponti et al., 2005). Also CD133 has been reported as surface marker to identify breast cancer-initiating cells (BCICs) (Wright et al., 2008), even if not extensively employed in BC. All of these subpopulations, obtained from pleural effusion or breast cancer cell lines (BCCLs), proved to be endowed with tumor-initiating ability when injected at low numbers (100-1,000 cells) in the mammary fat pad of immunocompromised mice (Al-Hajj et al., 2003; Ponti et al., 2005).

In addition to certain surface markers, the enzymatic activity of ALDH1 (Aldefluor^{TM} assay) also proved to identify a fraction of BCICs with higher tumorigenic ability compared to the remaining subpopulation (Ginestier et al., 2007). ALDH1+ cells show only partial overlap with the $CD44^{+}CD24^{low/-}$ population, displaying varying fractions (0.7-19%) among the different tumors (Angeloni et al., 2015).

The capability of efflux of the Hoechst 33342 or rhodamine-123 dyes represents another feature of stem-like cells and is used to identify the so called “side-population”, *i.e.* cells overexpressing the ATP-binding cassette transporters and with high tumorigenic ability. These cells account for a limited fraction of mammary epithelial cells and can be isolated also from clinical tumors and human BC cell lines (BCCLs; Patrawala et al., 2005).

Compared to normal breast epithelium, $CD44^{+}CD24^{low/-}$ BC cells showed a signature accounting for 186 differentially expressed genes, called invasiveness gene signature (IGS), with prognostic significance (Liu et al., 2007). It appears to be overexpressed in basal-like BC, likely ER , although its prognostic role is more evident within ER-positive BCs. Gene expression profile analysis including 76 of the 186 IGS genes and performed on BC specimens from a series of 110 postmenopausal patients treated with radical or conservative surgery plus radiotherapy and under adjuvant monotherapy with tamoxifen (Coradini et al., 2008) was able to discriminate cases with worst prognosis, whose tumor gene expression pattern showed a direct association to IGS, from disease-free cases, whose transcriptional profile was inversely associated to the IGS signature (Santilli et al., 2011). Moreover, the fraction of ALDH1+ BC cells correlates with basal-like and HER2-positive subtypes (Resetkova et al., 2010), and is predictive of clinical outcome in different BC clinical settings (Ginestier et al., 2007; Charafe-Jauffret et al., 2010).

Finally, another approach to enrich tumor cells with the BCIC subpopulation is the long-term treatment with chemotherapeutic agents, which takes advantage of the intrinsic chemoresistance of these cells (Calcagno et al., 2010; Li H.Z. et al., 2008).

Several signaling pathways essential for BCIC survival, self-renewal and tumorigenic properties are partially shared with those of normal mammary development, including Notch, Wnt and Hedgehog (Stylianou et al., 2006; Lamb et al., 2013; Fan et al., 2013). Beside canonic stemness pathways, the deregulation of PI3K/Akt/mTOR and JAK/STAT3 has been found to be implicated in BCIC maintenance and drug resistance (Zhou et al., 2007). Interplay between HER2 activation and Notch pathway has been recently postulated for HER2 overexpressing BC cells (Magnifico et al., 2009). Moreover, cytokines as IL-8 exert a stimulatory effect on BCICs via autocrine loop (Singh et al., 2013), and reciprocal interactions with mesenchymal stem cells via paracrine signaling involving CXCL7 and IL-6 has been demonstrated to create a suitable niche that preserve and fuel BCIC stemness capability (Liu et al., 2011).

Low proliferation rate and the expression of multi-drug resistance proteins represent two features of BCICs that have been linked to their intrinsic resistance to conventional anticancer therapies. CSCs may enter a quiescence status that protects them from antineoplastic agents that target dividing cells. Indeed, increased number of $CD44^{+}CD24^{low/-}$ cells was observed in the residual tumor following primary systemic therapy in BC patients (Li X. et al., 2008; Lee et al., 2011), but also after radiotherapy (Philips et al., 2006; Lagadec et al., 2010), suggesting a possible mechanism for BCIC selective growth and consequently relapse. Inducing de-differentiation in BCICs might render them more sensitive to standard therapies, taking advantage of their high level of plasticity that allow shifting from a quiescent to a more proliferative state (Liu et al., 2013), and some promising strategies, as the use of all-trans-retinoic acid and HDAC inhibitors, have been already proposed (reviewed in Angeloni et al., 2015).

More recently, among the huge number of models and mechanisms which have been proposed to explain resistance to antiestrogens (AEs) in BC, also tumor cells with stem-like properties were suggested as possible cause of tumor relapse after hormone therapies. BCIC models derived from the MCF7 cell line and propagated as mammospheres proved to have intrinsic resistance to AEs (Ao et al., 2011), also showing a down-regulation of ER-related genes even in basal conditions (Callari et al., 2016). In addition, short term exposure to AEs selectively enriches for BCICs via JAG1-NOTCH4 axis both *in vitro* and in patient-derived xenografts (Simões et al., 2015).

The first evidence for a link between CSCs and epithelial-to-mesenchymal transition (EMT) was reported in a seminal work showing that induction of EMT in immortalized human mammary epithelial cells was sufficient to induce the expression of stem cell markers; consistently, this was accompanied by increases in the formation of mammospheres, colonies in a soft agar assay and tumorigenicity in immunodeficient mice (Mani et al., 2008). More recently in normal breast tissue overexpression of the transcription factors Slug and Sox9 were enough to push luminal lineage cells into a more stem-like state, while only Sox9 was required in basal cells that already expressed the EMT associated transcription factor

Slug (Guo et al., 2012). Recently it was shown that $CD44^{+}CD24^{low/-}$ and ALDH1+ populations of BCSCs are plastic and have the capacity to transit between EMT and mesenchymal-to-epithelial (MET) transition states (Liu et al., 2013). The $CD44^{+}/CD24^{-}$ population displayed a molecular pattern of EMT, such as low expression of E-Cadherin and high levels of vimentin, and tended to be quiescent (EMT-CSCs). The ALDH-positive population, on the other hand, had a relatively opposite phenotype with high expression of E-Cadherin and low expression of vimentin, and was also much more proliferative, which pointed towards a more epithelial signature (MET-CSCs). The transition between these two states was postulated to be critical for tumor expansion (reviewed in Brabletz, 2012): while the EMT-CSCs sit at the invasive edge and allow the tumor to invade other tissues, the proliferative MET-CSCs likely drive tumor cell growth in the tumor inner part. When tumor conditions change or the invasive edge becomes the interior of the tumor, the two CSCs can change their status. This reversible, metastable epithelial-mesenchymal plasticity of BCICs is closely connected to current model of cancer metastasis postulating that EMT drives tumor cell dissemination and a consecutive MET drives metastatic colonization. In the case of BC, the $CD24^{-}CD44^{+}$ EMT-like BCSCs mediate tumor invasion toward the basal membrane and neighboring tissues and into the blood, where they survive due to their intrinsic quiescence and *anoikis* resistance. After extravasation in the circulation, these mesenchymal-like BCICs form micrometastases in distant organs, where metastatic niche or specific microenvironment in distal sites induces MET, which drives BCIC self-renewal and generation of macrometastases. An exception to this may be the claudin-low BCs which are characterized by a mesenchymal phenotype (Prat et al., 2010). These tumors may contain CSCs that simultaneously display EMT and MET properties. The existence of CSCs that are simultaneously invasive and proliferating might contribute to the very aggressive nature of this BC subtype.

Considering that CSCs may enter a dynamic EMT-MET state and that BCs are classifiable in different subtypes, each characterized by cells with more or less marked epithelial or mesenchymal phenotype, Wicha and colleagues summarized the features of

CSC subsets according to the BC molecular subtypes and proposed a bimodal model (Brooks et al., 2015; Figure 1.1.11). TNBCs can be ascribable to the molecular claudin-low or basal subtype. While the former is characterized by a high proportion of $CD44^{+}/CD24^{low/-}$ and ALDH1-positive CSCs as well as mesenchymal-like bulk tumor cells, the basal subtype, often associated with BRCA1 loss of function, contains epithelial bulk cells along with a subcomponent of mesenchymal bulk tumor cells and a higher proportion of epithelial ALDH1-positive CSCs. The CSCs subsets give rise to bulk tumor cells that are characterized by an epithelial morphology and lack of expression of ER or PgR. Also the HER2-positive subtype is characterized by a high proportion of ALDH1-positive CSCs that give rise to epithelial bulk populations, but in this case they may or may not express ER and PgR (Korkaya and Wicha, 2013). Luminal B breast cancers generally display a lower proportion of cells expressing CSC markers than either HER2-positive or TNBCs. The bulk cells in these tumors are highly epithelial in morphology and a proportion expresses ER and PgR. Luminal A tumors display the lowest proportion of cells expressing CSC markers and, accordingly, are also those with the best prognosis compared to the other subtypes. This model sustains the existence of an entirely distinct level of BC intratumor heterogeneity in which each of the above described subpopulations of cancer cells within a tumor contain its own fraction of BCICs and a bulk of non-BCICs that share the same genetic and epigenetic background (Pattabiraman and Weinberg, 2014). BCIC marker plasticity enables them to pass from a more proliferative epithelia-like state to a more quiescent, invasive and mesenchymal-like, each capable of generating their respective epithelial or mesenchymal bulk cell progeny which interact and reinforce the CSC one (May et al., 2011; Liu et al., 2013).

Single-cell sequencing approaches allowed the identification of BC cells exhibiting genomic diversity within the same tumor (Torres et al., 2007; Curtis et al., 2012; Wang et al., 2014). As carcinomas are the result of a succession and accumulation of genetic and epigenetic alterations, it seems plausible that CSCs are invariably present within incipient tumors at the various stages of progression, thus each of the cell populations formed at

intermediate stages of tumor evolution also should contain its own subpopulation of BCICs (Pattabiraman and Weinberg, 2014). However, this model is in contrast with the stem cell low mitotic rate, which represents one of their main features.

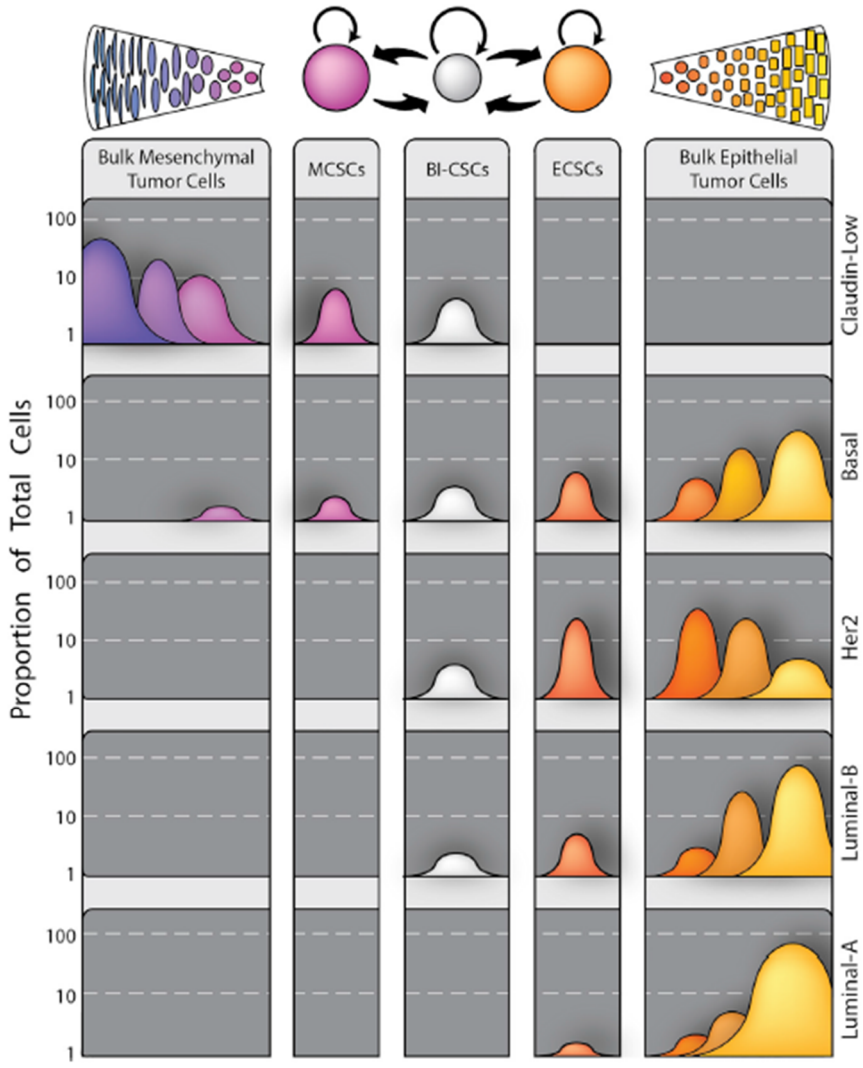


Figure 1.1.11. Scheme of a model illustrating CSCs with different phenotype across the spectrum of BC molecular subtypes. The different molecular subtypes of BC are characterized by varying proportions of CSCs in complete mesenchymal (MCSCs) or epithelial (ECSCs) state as well as CSCs in transient state from epithelial to mesenchymal state and vice versa (BI-CSCs), that give rise to the bulk of differentiated epithelial or mesenchymal progeny. *Brooks et al., 2015*

Despite the diversity of genetic changes driving the different molecular subtypes of BC (Cancer Genome Atlas Network, 2012), BCICs within these subtypes proved to have similar patterns of gene expression (Liu et al., 2013) and to share similar genomic alterations with the bulk of non-BCICs (Klevebring et al., 2014), suggesting that genetic diversity cannot

account for BCIC heterogeneity. Consistently with this observation, differentiated mammary epithelial cells were described to be able to spontaneously acquire stem cell features and to restore the equilibrium of the population of origin (Chaffer et al., 2011; Gupta et al., 2011), thus suggesting that, in line with CSC plasticity and in contrast with the hierarchy model, BCICs may arise *de novo* from non-BCICs.

1.2. CIRCULATING TUMOR CELLS

1.2.1. Biological properties of circulating tumor cells

Circulating tumor cells (CTCs) are by definition tumor cells that are detected while in transit within the blood stream (Yu et al., 2011). CTCs originate from the primary tumor (PT) and represent a subpopulation of tumor cells which acquired traits of invasiveness and motility, and which have the ability to intravasate and survive in blood for short or long periods in the attempt to re-establish secondary tumors in other sites (Massagué and Obenauf, 2016). However, this classical view of the metastatic cascade as an unidirectional process has been recently overturned by studies with experimental models, which demonstrated that CTCs may re-infiltrate an established tumor, enriching it with aggressive cells after a period of dissemination, a process called “tumor self-seeding” (Norton and Massagué, 2006; Kim et al., 2009) and which might have important clinical implications in terms of loco-regional recurrence (Bidard et al., 2009).

Metastases generally appear in a late phase of tumorigenesis, when a PT achieves a critical mass of cells and acquires an aggressive phenotype, according to the linear model of tumor progression, but dissemination can also represents an early independent event, as proposed in the parallel progression model (Klein, 2009) and as frequently observed in BC. In fact, tumor cells can disseminate systemically at the earliest stage of mammary epithelial cell transformation (Hüsemann et al., 2008), even before they become chromosomally instable (Schardt et al., 2005), and they can be detected also in the blood of patients with node-negative operable BC (Xenidis et al., 2006), or in patients with atypia (FEA) or DCIS (Franken et al., 2012).

Despite the marked proficiency of BC cells in disseminating, not all CTCs are able to give rise to overt metastases since their survival in blood is hampered by a number of obstacles, as fluid shear stress exerted by blood flow (Mitchell and King, 2013), *anoikis* (Kim

et al., 2012), and the attacks of the immune system (CD47; Mohme et al., 2016). Indeed, consistently with the notorious scarce efficiency of the metastatic process (Luzzi et al., 1998), CTC half-life was estimated to be short, from 7-9 minutes to 2.4 hours, according to findings reported in experimental and clinical studies (Fidler, 1970; Sasportas and Gambhir, 2014; Meng et al., 2004b). Therefore, on the basis of these observations, the existence of a subset of CTCs with metastasis-initiating properties has been hypothesized.

The hypothesis that only a subpopulation of CTCs is really capable of reinitiating macroscopic tumor growth in a distant tissue has been reviewed and addressed by various authors (Oskarsson et al., 2014; Celià-Terrassa and Kang, 2016). The first evidence that CTCs isolated from BC patients are able to give rise to metastases in experimental models was provided in 2013 (Baccelli et al., 2013). In this study the authors transplanted CTCs isolated from metastatic BC patients into the femoral medullary cavity of immunocompromised mice and demonstrated that the metastatic potential of the bulk of CTCs obtained after depletion of hematopoietic cells was negligible when less than 1,000 were injected. In contrast, mice receiving higher number of CTCs enriched in a subpopulation of cells with $EpCAM^{+}CD44^{+}CD47^{+}MET^{+/-}$ phenotype developed multiple lesions at bones, lungs and liver. Subsequent to this study, other subpopulations of metastasis-initiating disseminated cells were identified. CTCs can be found in circulation as single cells or clusters of cells, in different kinds of neoplasms (Brandt et al., 1996; Molnar et al., 2001; Kats-Ugurlu et al., 2009; Hou et al., 2012), including BC (Stott et al., 2010; Yu et al., 2013). The seminal work by Aceto and colleagues (Aceto et al., 2014) revealed that CTC clusters, despite representing a rare subpopulation of CTCs (about 2%), display 23- to 50-fold higher metastatic potential compared to the single CTC subset. Interestingly, CTC clusters have proven to have oligoclonal origin and to represent the result of cell aggregation events, modulated by the junction protein plakoglobin, which take place before and not after intravasation. A more recent work corroborated these results in a different experimental model and also demonstrated that clusters are composed by two molecularly distinct subpopulations whose proportions vary during metastasis, as tumor cells at the leading edge

of the invading cluster in the PT and clusters circulating in blood are 20-fold enriched in KRT14+ cells compared to the PT and pulmonary macrometastases (Cheung et al., 2016).

These results seem to be apparently in contrast when considering the size of epithelial tumor cells, (20 to 30 µm diameter), and the size of capillary bed (about 8 µm diameter). Therefore, the ability of CTC clusters to pass through capillaries was questioned. A recent technical paper provided an explanation for such an aspect of tumor dissemination process by using CTC clusters isolated from clinical samples, microscale devices, computational simulations and experimental animal models (Au et al., 2016). The authors could show that over 90% of clusters containing up to 20 cells did successfully traverse 5- to 10-µm diameter constrictions because of rapid and reversible reorganization of cells into single-file chain-like geometries. This data highlights once more the importance of CTC morphological plasticity, which reduces their hydrodynamic resistances while passing through capillaries and, importantly, allowes the CTC cluster to re-establish the original rounded morphology and maintain its viability and function.

Plasticity and heterogeneity have been frequently observed in CTCs also at molecular level (Figure 1.2.1). Notoriously, carcinoma cells activate a trans-differentiation program, called epithelial-to-mesenchymal transition (EMT), which endows cells with migratory and invasive properties, induces the acquisition of stem cell traits, prevents apoptosis and senescence, and contributes to immunosuppression and drug resistance (Thiery et al., 2009; Brabletz, 2012). Basically, key molecular players, such as Twist, Snail, Slug and others, activate a transcriptional program which provokes the down-regulation of epithelial genes, as EpCAM, claudins, occludins, E-cadherin etc., and the up-regulation of mesenchymal genes, as vimentin, fibronectin, N-cadherin, etc. (Sleeman and Thiery, 2011). As a consequence, cells undergo crucial changes in their cytoskeleton and lose the expression of adhesion molecules, two events that confer the typical spindle-shaped morphology and culminate in increased cell motility (Thiery et al., 2009). The shift toward a different phenotype is characterized by a sequence of events during which cells can temporarily transit in an intermediate state and present both epithelial and mesenchymal features or display a partial

EMT-phenotype (Mego et al., 2010; Nieto et al., 2016). Despite technical difficulties in describing this transitory process at CTC level, both in animal models and human cancers, experiments with a dynamic *in vivo* model obtained from xenograft of MDA-MB-468 cells showed that EMT occurs at the PT level and is responsible for the generation of CTCs (Bonnomet et al., 2012). Importantly, longitudinal monitoring of CTCs isolated from BC patients during systemic therapy provided evidence for the first time for the co-existence of cells with either complete epithelial or transient epithelial-mesenchymal or complete mesenchymal CTCs, at different proportion and in the same patient, according to the therapy response status (Yu et al., 2013). In particular, the fraction of cells with marked mesenchymal features was preponderant during tumor progression compared to stages during which patients had reached good response to treatment, consistently with the model attributing therapy resistance traits to cells with EMT-phenotype and with the concept of the EMT process as a continuum (Mego et al., 2010; Nieto et al., 2016).

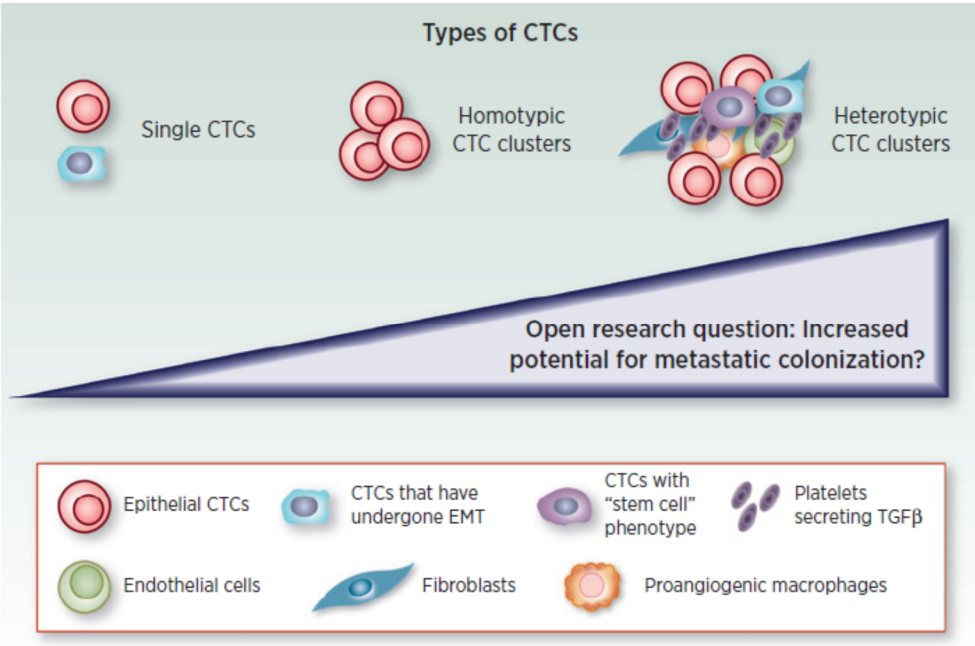


Figure 1.2.1. The CTC population is heterogeneous. CTCs may shift from an epithelial to a mesenchymal phenotype, and can transit in blood as single cells or clusters, either forming homotypic interactions with other tumor cells or heterotypic interactions with other kinds of cells. *Ignatiadis et al., 2015*

If EMT is a requisite for carcinoma cells to become CTCs, the return from a mesenchymal to an epithelial phenotype (called mesenchymal-to-epithelial transition, MET) seems to be the necessary driving force for CTC homing and efficient colonization. In agreement with this view, comparison of central areas of primary colorectal carcinomas and matched metastases revealed the presence of similar patterns related to differentiated epithelial cells (Brabletz et al., 2001). Two more recent papers elegantly demonstrated that down-modulation of Twist1 and Prrx1 EMT transcription factors and subsequent re-differentiation (MET) are necessary for distant colonization (Ocaña et al., 2012; Tsai et al., 2012). However, the role of EMT in dissemination is not completely understood and data on the involvement of this process in metastasis are still controversial. Indeed, recent studies with a lineage-tracing transgenic BC model showed that inhibiting EMT by overexpressing the microRNA miR-200 did not affect lung metastasis development (Fischer et al., 2015).

Preclinical studies have demonstrated that CTC clusters are not solely formed by tumor cells but also include accessory cells, as fibroblasts, leukocytes, endothelial cells, pericytes and platelets, which form heterotypic interactions with CTCs. Duda and colleagues demonstrated that fibroblasts are shed from the PT with CTCs and confer them increased survival and early growth advantage at metastatic site; CTCs, in turn, are likely to activate stromal cells to a CAF-state as the majority of fibroblasts carried over by CTCs were αSMA+ and FSP1+ (Duda et al., 2010).

It is widely recognized that tumor-associated macrophages (TAMs) facilitate almost every step of the metastatic cascade, including intravasation, dissemination and establishment of metastasis at secondary sites (Belgiovine et al., 2016). Multiphoton imaging allowed the visualization of perivascular macrophages during tumor cell intravasation (Wyckoff et al., 2007). More strikingly, studies with clodronate liposomes and zoledronic acid showed that elimination of macrophages significantly influenced CTC numbers and tumor metastasis formation (Rodrigues et al., 2011; Zhang et al., 2010). Beside the involvement of macrophages in EMT (Su S. et al., 2014), an alternative model considers fusion of monocyte-derived cells with tumor cells in blood (macrophage-tumor cell fusions, MTFs) as a

possible process to promote colonization (Pawelek and Chakraborty, 2008; Lazova et al., 2013; Clawson et al., 2015). Interestingly, circulating giant macrophages-like cells were found bound to CTCs in patients with different kinds of tumors, including BC (Adams D.L. et al., 2014; Mu et al., 2016).

CTCs are also able to attract a multitude of aggregating platelets (Camerer et al., 2004; Nieswandt et al., 1999), which in turn protect tumor cells via TGF-β signaling and stimulate the acquisition of aggressive phenotype, facilitating the metastatic process (Gay and Felding-Habermann, 2011; Labelle et al., 2011).

Other allies of CTCs are represented by tumor-associated endothelial cells, which protect them from *anoikis* (Yadav et al., 2015) and increase the number of CTCs and spontaneous metastases in experimental models (Stacer et al., 2016). Endosialin-expressing pericytes, too, have proven to promote tumor cell intravasation in a cell contact-dependent manner, resulting in elevated numbers of CTCs in mouse models (Viski et al., 2016). Furthermore, circulating tumor-associated neutrophils (cTANs) contribute to CTC survival by suppressing peripheral leukocyte activation (Zhang et al., 2016).

1.2.2. CTC detection and clinical applicability of CTC blood-tests

CTCs offer an opportunity to obtain key information for the development of personalized medicine. However, since these cells represent rare events, due to the presence of circulating cell populations of hematopoietic origin (few million white blood cells and a billion red blood cells per milliliter of blood), and are heterogeneous and subject to changes during the various phases of dissemination or in response to therapy, it is clear that their value as biomarkers strongly depends on the detection method. Operational definition of CTCs is strictly related to the methodological approach as no specific marker exists to discriminate them from other cells. No single parameter, such as size, morphology, protein expression or chromosomal aberration pattern can be considered sufficient to define a CTC.

This implies that pre-enrichment and detection methods should be combined to increase specificity, although at the cost of sensitivity.

CTCs were observed for the first time in 1869 in the blood of a man with metastatic cancer by Thomas Ashworth (Ashworth, 1869). He found cells with identical aspects of those present in solid malignant lesions and concluded that “One thing is certain, that if they [CTC] came from an existing cancer structure, they must have passed through the greater part of the circulatory system to have arrived at the internal saphena vein of the sound leg”. Since then, many decades have passed until CTCs could be systematically studied in human samples.

In more recent years, driven by technical progress, numerous assays and devices to isolate and detect CTCs have been described. The most commonly used approaches for CTC analysis were briefly described in Table 1, including positive and negative aspects and key methodological papers, with special reference to BC studies.

Blood samples can be enriched for CTCs by positive selection approaches, generally intended as solid-phase capture strategies, as magnetic beads or devices coated with antibodies directed against epithelial and tumor surface markers. Among them, the epithelial marker EpCAM has been widely exploited to catch CTCs, but this approach may underestimate the number and composition of CTCs since it fails to detect tumor cells with mesenchymal features, which might provide further and probably more useful information on therapy response (Wicha and Hayes, 2011). Therefore, not only technical but also biological specificity should be taken into account as the detected cancer cell might represent only a fraction of the entire CTC population or might possess variable ability to give rise to metastases. Alternatively, CTCs can be isolated taking advantage of their size and morphology (Vona et al., 2000), by using devices which allow physical size-based exclusion of hematopoietic cells and unbiased CTC isolation with regard to the expression of surface markers. Basically, tumor cell diameter is around 20-30 µm, much larger compared to the vast majority of hematopoietic cell populations. In this perspective, specific devices have been created in order to entrap CTCs, but also other atypical cells, onto microporous

membranes (Table 1). Despite this approach appears free from technical limitations, the blood flow rate has proven to be crucial for isolation of intact cells, especially for CTC clusters (Stott et al., 2010), whose biological and clinical relevance has been demonstrated both in experimental and clinical studies.

Semi-automated high-throughput technologies have been recently introduced, determining a revolution in CTC studies (Table 1). The DEPArray™ technology is based on the ability of a non-uniform electric field to exert forces on neutral, polarizable particles, such as cells, that are suspended in a liquid (Medoro et al., 2003). This electrokinetic principle, called dielectrophoresis (DEP), can be used to gently trap and move cells within DEP “cages”, allowing image-based isolation of single cells. However, this system does not work on whole blood volumes, but on few microliters of sample. Therefore, a preliminary enrichment step is required. Another powerful technology is represented by the CTC-chip, which is based on the generation of a microfluidic vortex that increases the chance of interaction between target cells and antibodies on the surface of a herringbone-like device (Nagrath et al., 2007). Differently from the DEPAarray^{TM}, this technology is suitable for isolation of both single CTCs and CTC clusters.

A second phase necessarily follows the pre-enrichment step as, once isolated, the presence of CTCs has to be verified and a multitude of approaches can be applied to characterize them. CTCs can be directly counted and characterized by *in situ* analyses, cultured for functional studies or lysed for nucleic acid extraction and molecular analysis (Yu et al., 2011). Cytological staining protocols allow direct count and morphological analysis of CTCs according to classical cytopathological criteria for identification of malignant cells (Hofman et al., 2011). This approach requires that sample analysis should be performed by a pathologist, although elements subjected to doubt interpretation, as cells having nonhematologic origin but with uncertain malignant or benign feature, can be observed (Hofman et al., 2011).

Table 1. Methodological approaches for CTC analysis					
Technology	**Enrichment method**	**Detection method**	**Pros**	**Cons**	**Ref.**
AdnaTest	CTC capture by immunomagnetic beads (wash buffer for leukocyte contamination reduction).	Multiplex semi-quantitative PCR.	High sensitivity (2 CTCs in 5 ml of blood) due to the detection method. Reliable CTC gene profiling feasible by low-density and high-density arrays.	Enrichment dependent on the expression of surface markers (EpCAM, MUC1, ERBB2, EGFR). Direct CTC count not feasible.	Tewes et al., 2009; Aktas et al., 2009
CAM assay	Depletion of leukocytes or FACS.	Cell culture on a cell-adhesion (collagen) matrix and detection of CTCs with invasiveness properties.	Unbiased enrichment. CTC culture. Cell viability. CTC analysis at functional level.	Possible modifications of CTC molecular features due to culture conditions.	Lu et al., 2010
CellSearch® system	CTC capture by immunomagnetic beads.	Immunofluorescence-based detection by KRTs and CD45 staining. Nucleic acid analysis.	Direct CTC count and visualization. Semi-automated. Approved by FDA.	Enrichment dependent on the expression of EpCAM on the cell surface. Maximum 4 fluorescence channels for CTC detection.	Riethdorf et al., 2007
CellSieve™ microfilters	Size-based isolation by blood filtration through microporous membranes (pores diameters 7 µm).	Direct visualization and *in situ* analyses, CTC culture or nucleic acids analysis.	Precision, uniform and high porosity microfilters. Filter flatness facilitates images acquisition. High recovery rate. Unbiased CTC enrichment. Identification of single CTCs, CTC clusters and other atypical cells.	No possibility to retrieve intact isolated cells.	Adams D.L. et al., 2014
Microfluidic HB-chip	No pre-enrichment required. Highthroughput device which generates microvortices, increasing the number of interactions between CTCs and an antibody-coated surface.	Imaging (histopathological stains and IF).	Minimal damage of sensitive CTC populations. Coating with different kinds of antibodies. Good capture efficiency.	Enrichment dependent on the expression of specific markers. Manual pick-up of isolated cells for further analysis. Not commercially available.	Nagrath et al., 2007

Table 1. *(continued)* Methodological approaches for CTC analysis					
Technology	**Enrichment method**	**Detection method**	**Pros**	**Cons**	**Ref.**
EPISPOT assay	Depletion of leukocytes.	Cell culture and IF-based detection of proteins secreted by CTCs.	Unbiased enrichment. CTC culture. Assessment of CTC viability and biological activity (secretion of PSA, MUC1, FGF2, etc.).	Possible modifications of CTC molecular features due to culture conditions.	Ramirez et al., 2014
DEPArray™	Leukocyte depletion with other devices.	Image-based single cell sorting by dielectrophoresis of IF-labeled cells.	Pure and single cell isolation and characterization.	30% cell loss due to the technical limits of the instrument. Only 4 immunofluorescence channels. Limited number of input (≤40,000) cells per chip.	Peeters et al., 2013; Fernandez et al., 2014
Ikoniscope®	Density gradient centrifugation or filtration and CTC labeling with immunofluorescent antibodies (KRTs, EpCAM).	FISH analysis by robotic fluorescence microscope imaging system.	Reliable CTC identification by analysis of chromosomal alterations.	Tumors with known chromosomal rearrangements. Probes for specific FISH tests required.	Ntouroupi et al., 2008
ScreenCell®	Size-based isolation by blood filtration through microporous membranes (pores diameters ~8 µm).	Direct visualization by *in situ* analyses. CTC culture or nucleic acids analysis.	Direct CTC count, morphological/cytological evaluation and immunological analysis (ICC and IF). High sensitivity (1 CTC in 3 ml of blood). Unbiased CTC enrichment. Identification of single CTCs, CTC clusters and other atypical cells.	Micro-coagulation may cause blood flow arrest during filtration. Fast filtration rate (~1 ml blood/min) may damage CTC clusters.	Desitter et al., 2011; Mu et al., 2016
OBP-401 assay	No pre-enrichment required. Infection with telomerase-specific replication-selective adenovirus.	GFP-positive cells visualization. FISH analysis.	High specificity for tumor cells due to enhanced virus replication compared to normal cells.	Not applicable to hTERT-independent neoplasms.	Kim et al., 2011

Protein analysis by immunocytochemistry (ICC) and immunofluorescence (IF) may provide further information, but subsets of CTCs can be missed due to their heterogeneity and lack of expression of some markers. Moreover, atypical cells called macrophages-like cells were found in the blood of cancer patients, identified as giant cells expressing monocytic markers, but in some cases either expressing or not the common leukocyte antigen CD45 and also resulting positive for epithelial and endothelial protein markers (Adams D.L. et al., 2014). Therefore, CTC identification by direct staining for proteins is limited by the complexity and variety cell types circulating in blood and lack of univocal markers.

Molecular analysis of CTCs, following positive or negative selection approaches, can be performed by *in situ* techniques, as FISH or mRNA probe hybridization, or after cell lysis and nucleic acids extraction. In the first case, direct visualization of target cells might help identifying them and provide reliable results, although limited number of assays can be used on the same sample. In the second case, extensive molecular analysis can be performed using high- or low-density arrays, qPCR, RNA sequencing, digital PCR and whole exome or whole genome sequencing. Nevertheless, in this situation the purity of the sample has strong implications on the type of markers that can be analyzed at mRNA level. For example, genes characterizing mesenchymal and stem-like cells, such as *VIM* and *ALDH1*, are also expressed by leukocytes. Parallel analyses on healthy donor case series or on patients' blood samples after CTC depletion (for example by incubation with mock-functionalized antibodies with the same isotype of those used for CTC capture) are necessary to identify reliable positivity threshold values to assign CTC positivity status or CTC-marker positivity. Actually, the scenario is more complex than expected as there is evidence that marker genes commonly used to detect CTCs in PCR-based tests, as squamous-cell carcinoma antigen, epidermal growth factor receptor, mammaglobin, small breast epithelial mucin, are expressed in normal, mitogen-stimulated peripheral blood mononuclear cells (PBMNC). Thus, considering the inflammatory reactions often accompanying cancer development, but also other pathologies as inflammatory bowel

disease, molecular markers were suggested to be validated not against normal peripheral blood cells, but against activated lymphoid cells, such as *in vitro* mitogen-stimulated PBMNC, in order to reduce false-positives (Kowalewska et al., 2006).

The opportunity to characterize single cells completely resets the need for specificity tests as the nature of the target cell can be verified exploiting one of the most important hallmarks of cancer, which is represented by genetic instability. Basically, DNA ploidy status and copy number variation analyses definitively allow target cell rating as real CTC or not. Moreover, mutational analysis does not need additional specificity tests beyond those required for technical threshold definition, since the investigated genetic alteration should be private, *i.e.* of the tumor, and theoretically not detectable at germline level.

CTCs hold the key to understand the biology of metastasis and provide a biomarker to noninvasively measure the evolution of the tumor during treatment as also disease progression. Indeed, differently from cell-free circulating nucleic acids and proteins, CTCs are biological entities whose clinical significance as biomarkers can be assessed at different levels. First, CTC presence or absence in the sample analyzed can be informative *per se* on disease aggressiveness and response to therapy. In non-metastatic BC detection of at least 1 CTC in 7.5 mL of blood or positivity for CK-19 assessed by RT-PCR are sufficient to identify patients at high risk of relapse (Saloustros et al., 2011; Lucci et al., 2012). Interestingly, longitudinal monitoring of CTC status is likely to be much more informative than CTC assessment at a single time point (Musella et al., 2015; Fina et al., 2016). Second, the number of detected CTCs in a defined volume of blood generally mirrors the tumor burden, thus indicating that prognostic cut-off need to be identified to predict the clinical outcome (Cristofanilli et al., 2004). Finally, CTC characterization may outperform simple CTC enumeration or CTC detection by EpCAM-based methods, as already shown in BC and CRC studies (Mostert et al., 2015; Gazzaniga et al., 2011).

At present, technical constraints limit the applicability of CTC-test in the clinical context (Alix-Panabières and Pantel, 2014), except for the FDA-cleared technology called CellSearch^{TM} system (Table 1), approved for CTC monitoring in metastatic breast, colorectal

and prostate cancers (https://www.cellsearchctc.com/clinical-applications/clinical-relevance). Nevertheless, CTC studies represent an opportunity that should not be missed and, even if hard to find, CTCs cannot be ignored and deserve further investigations.

The route toward personalized medicine is long, but certainly standardization studies are required to make CTCs a useful biomarker that could support clinical decision-making.

1.2.3. Clinical significance of circulating tumor cell count in breast cancer

CTC-based tests may rely on CTC detection by applying cyto-morphological criteria or by the analysis of the expression pattern of specific proteins to distinguish tumor cells from other cells transiting in blood. The majority of platforms currently available holds great promise to improve CTC detection and monitoring in the clinical context. In the last decade the assessment of CTC levels has been widely tested as a new prognostic tool in BC. In 2004, CTC enumeration using the CellSearch system (Janssen Diagnostics, Raritan, NJ, USA) was shown to be significantly associated with progression-free survival (PFS) and overall survival (OS) in 177 patients with metastatic breast cancer (Cristofanilli et al., 2004). The hazard ratio (HR) for the difference between late and early progression of disease reached a plateau at 5 CTCs per 7.5 mL or higher. In the same cohort, changes in CTC count after the initiation of a new line of therapy were also shown to correlate with PFS and OS (Cristofanilli et al., 2005). Since then, numerous observational studies on the prognostic and predictive value of CTC count have been reported (reviewed in Bidard et al., 2016). The prognostic value of CTC levels was also observed during treatment, as by combining dichotomized CTC count (high or low) at two time-points (baseline and after one cycle of treatment), four different PFS profiles were obtained: the worse prognosis was seen in patients with high CTC count at both time points, as expected, while patients with a high CTC count at baseline but a low CTC count after one cycle of therapy had a much better prognosis, almost similar to that of patients with low CTC count at baseline, thus demonstrating that CTC count may also represent a dynamic prognostic marker (Hayes et

al., 2006). Moreover, CTC count was shown to accurately anticipate the information provided by radiological imaging since patients without disease progression according to instrumental response and with ≥5 CTCs had a shorter median overall survival than those with lower CTC number; on the contrary, patients who presented with radiological progression and <5 CTCs had a longer survival than those with elevated CTCs (Budd et al., 2006).

In 2014 a pooled analysis of 1944 individual patient data provided level-1 evidence for the clinical validity of baseline CTC ≥5 per 7.5 mL of blood as an independent marker of poor prognosis in metastatic BC (Bidard et al., 2014). Hazard ratio of survival between high and low CTC counts increased together with the threshold used to define high CTC count. Finally, in contrast to serum tumor markers, such as CEA and CA15.3, which showed inferiority to CTC count, the study demonstrated that by adding CTC count and its change during therapy to an optimized clinico-pathological model, the prognostic value of the model significantly increased (Bidard et al., 2014).

However, the value of CTC enumeration for treatment decision making in metastatic BC was not validated. It was prospectively tested in the Southwest Oncology Group (SWOG) S0500 clinical trial (Smerage et al., 2014), in order to evaluate the benefit of an early change in chemotherapy for patients with persistent high CTC level at first follow-up after starting first-line chemotherapy. Although CTCs were strongly prognostic, an early switch to an alternative chemotherapy did not increase overall survival. These negative results were supposed to be related to the study design, since switching from one ineffective therapy to another ineffective therapy it is expected not to change the outcome even if this shift is anticipated thanks to the information given by the CTC status, and changing treatment based on CTC molecular features now seems to be a more promising approach.

Another issue related to the failure of the first CTC-based clinical trial is represented by the criteria used to define the CTC threshold. The 5 CTC threshold was the median CTC count that maximized the significance of the log-rank test in the first paper by Cristofanilli (Cristofanilli et al., 2004). Further data, obtained in the large European pooled analysis of CTC in metastatic BC patients, showed that the survival hazard-ratio increases together with

the CTC count, with no clear threshold (Bidard et al., 2014). Furthermore, it is critical to point out that the ≥5 CTC/7.5 mL threshold is an overall prognostic factor in metastatic BC of global survival, but was never optimized to detect early tumor progression, *i.e.* resistance to chemotherapy with tumor progression within the first 3-4 months; therefore, Bidard and colleagues recommend that thresholds for clinical validity should be distinguished, whenever needed, from those intended for clinical utility (Bidard et al., 2016).

At present two main studies are ongoing and results are expected in the following two years. The CirCe01 trial (NCT01349842) is based on the early changes of CTC count, but patients are enrolled before the start of third line of chemotherapy and followed with the CTC test throughout the successive lines of chemotherapy. The efficacy of therapy is determined by comparison of CTC level before the first injection of each new line of chemotherapy with conventional clinical and radiological evaluation. Instead, the STIC CTC trial (NCT01710605) investigates the clinical utility of the prognostic value of baseline CTC count. In this trial, the choice of the first line of treatment (hormone therapy or chemotherapy) for relapsing hormone-positive BCs is determined either by the clinician or by the baseline CTC count: hormone-therapy if <5 CTC/7.5mL (CellSearch technique) or chemotherapy if =5. The main medical objective is to demonstrate the non-inferiority of the CTC-based strategy for the progression-free survival.

The authors of the first report on the phase III CirCe01 trial observed that among patients with ≥5 CTC/7.5 mL at baseline, a composite criteria (<5 CTC/7.5 mL or relative decrease ≥-70% of the baseline CTC count) showed better prognostication for early tumor progression (within 4 months) than the proposed <5 CTC/7.5 mL threshold (Helissey et al., 2015). The role played by this threshold-related issue in the SWOG0500 trial failure is potentially critical, as using ≥5 CTC/7.5 mL after 3-5 weeks on treatment as a predictive test of early (<6 months) tumor progression, only 60% of patients with a positive test (*i.e.*, ≥5 CTC/7.5 ml) have a PFS shorter than 6 months, while 75% of patients with <5 CTC/7.5 mL after 3-5 weeks on treatment have a PFS >6 months. The positive and

negative predictive values of the 5 CTC/7.5 mL threshold are therefore limited when used as a test to predict early tumor progression (Figure 1.2.2).

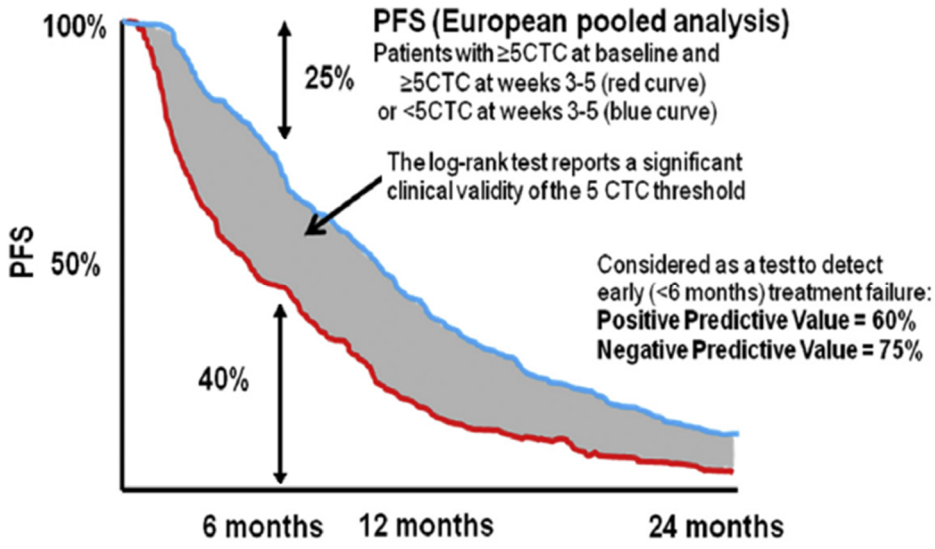


Figure 1.2.2. CTC count thresholds with clinical validity may have limited clinical utility. Progression-free survival (PFS) curves of metastatic BC patients with ≥5 CTC/7.5 mL at baseline, according the CTC count at weeks 3-5: the red curve displays the PFS of patients who retains a CTC count ≥5 CTC/7.5 mL after 3-5 weeks on treatment and the blue curve displays the PFS of patients in whom the CTC count decreases <5 CTC/7.5 mL after 3-5 weeks on treatment (curves inspired from the European pooled analysis (Bidard et al., 2014)). *Adapted from Bidard et al., 2016*

The use of CTCs as a prognostic and predictive marker was proven to be important also for patients with early BC, as preliminary prospective clinical trials suggest that the presence of CTCs at the time of primary diagnosis could predict early disease recurrence and reduced survival (Franken et al., 2012; Ignatiadis et al., 2008; Lucci et al., 2012; Pierga et al., 2008; Rack et al., 2014). Strikingly, in opposition to most of the currently used clinic-pathological prognostic factors, CTC detection studies using the CellSearch technique showed no association with the response to therapy, measured by the completion of a pathological complete response in the neoadjuvant setting. A pooled analysis including individual data from 3,173 patients with nonmetastatic (stage I–III) breast cancers confirmed that baseline detection of ≥1 CTC per 7.5 mL of blood (using the CellSearch system) is an independent predictor of poor disease-free, overall, breast cancer–specific, and distant

disease-free survival in both the neoadjuvant and adjuvant setting (Janni et al., 2016). These results fostered the initiation of the ongoing Treat CTC trial (NCT01548677), which assesses the efficacy of Trastuzumab to treat HER2-negative primary BCs with detectable CTCs. According to Treat CTC study design, patients with HER2-negative primary tumor who have completed either adjuvant chemotherapy (adjuvant cohort) or neoadjuvant chemotherapy and have residual disease in either the breast or the lymph nodes (neoadjuvant cohort) will be screened for the presence of CTCs using the CellSearch technology. Then, patients with detectable CTCs will be grouped to evaluate the effect of six cycles of therapy with Herceptin in CTC detection rate versus an observation untreated arm. The first report was published by Ignatiadis and colleagues (Ignatiadis et al., 2016), who described preliminary results obtained after screening 350 patients. According to CTC test results, performed in three central laboratories, 39 out of 318 (12%) assessable cases had ≥1 CTC/15 mL of blood after completing chemotherapy, data which is consistent with CTC detection rates already observed in non-metastatic breast cancers when using the CellSearch technology. Twenty-six out of 39 CTC+ patients, 20 (76%) with 1 CTC/15 mL of blood and 6 (24%) with 2 to 14 CTC/15 mL, were then assigned either to the Herceptin treated or untreated group; results of correlation analyses with the clinical outcome will be available in the following 3 years.

Interestingly, in the first stage of Treat CTC Trial, 1 out of 26 CTC-positive patients had 1 HER2-positive CTC by CellSearch-based score (Ignatiadis et al., 2016). Discordance in HER2 status between the PT and disseminated cells has been reported in many studies, although the real clinical efficacy of anti-HER2 therapies, either in HER2-positive or HER2-negative CTCs, is still controversial (Houssami et al., 2011; Meng et al., 2004a; Pestrin et al., 2012; Janni et al., 2016). A further level of complication is represented by recent evidence that HER2-positive and HER2−negative CTCs interconvert spontaneously, with cells of one phenotype producing daughters of the opposite within four cell doublings (Jordan et al., 2016). Moreover, the authors of these studies showed that although HER2-positive and HER2−negative CTCs have comparable tumor initiating potential,

differential proliferation favors the HER2-positive state, while oxidative stress or cytotoxic chemotherapy enhance transition to the HER2-negative phenotype.

1.2.4. Molecular characterization of circulating tumor cells in breast cancer

CTCs represent a unique opportunity to repeatedly monitor the metastatic process in patients and to study the molecular (genomic, transcriptomic and proteomic) processes associated with the spread of tumor cells. This opportunity has important implications for the development of blood-tests to monitor the expression of biomarkers in CTCs and possibly to support clinical decision-making.

Molecular characterization can be performed on single CTCs or on the CTC bulk. Both approaches present advantages and pitfalls. Single-cell analysis requires sophisticated technologies to isolate pure single CTCs and to enable sensitive and reliable detection of transcripts or genomic alterations. On the contrary, studies on the whole CTC population allow easier and faster recover of CTCs but the quality of the clinical sample depends on the CTC/background leukocytes ratio and requires the identification of reliable positivity cut-offs; moreover, these approaches do not allow exploring the CTC heterogeneity. However, independently from the method, CTC characterization was proven to outperform the simple CTC count in various studies.

Mostert and colleagues (Mostert et al., 2015) showed that a 16-gene expression signature of CTCs correlates with prognosis, in addition to, but also irrespective of, CTC count. In fact, among patients with <5 CTCs, who would be classified as good prognosis according to their CTC count by CellSearch, the 16-gene CTC profile was able to distinguish a good from an intermediate prognosis group, providing additional information to CTC count. However, the results were not validated in a further cohort, suggesting that the CTC isolation approach, which was EpCAM-based, and the panel of genes chosen underestimated the biological heterogeneity of CTCs.

Interestingly, many studies based on single CTC analysis are reporting high discrepancy between CTCs and matched primary and secondary lesions, as also among the same CTCs. *PIK3CA* mutation status, for example, was shown to be heterogeneous among tumor cells of different origin or within the CTC population, as also to change over time, according to disease progression (Pestrin et al., 2015; Deng et al., 2014; Markou et al., 2014). However, it is unclear whether these high discrepancies are due to either the limited sensitivity of single cell characterization techniques, leading to false negative results, or to a real intercellular heterogeneity, previously ignored.

Notwithstanding technical and interpretative difficulties, and the actual need for extensive and reproducible studies, CTC characterization deserves attention and effort to turn this great opportunity into readily usable tests and new therapeutic protocols.

1.3. THE BIOLOGY OF BREAST CANCER METASTASIS

1.3.1. Landmark discoveries and theories on cancer metastasis

The term “metastasis” was first applied to all maladies that seem to transpose from their point of origin to another organ with no specific reference to or knowledge of cancer. The first recorded definition of metastasis in the context of cancer diagnosis and treatment was made by Jean Claude Recamier in 1829 (Récamier, 1829) and since then metastasis was recognized as a multistage process during which malignant cells spread from the primary tumor to distant organs (Dorland, 1965; Christofori, 2006; Gupta and Massagué, 2006; Figure 1.3.1). After initial neoplastic transformation events, the establishment of a vascular network and local invasion of the host stroma occur as a consequence of tumor cell adherence loss, proteolytic degradation of the surrounding extracellular matrix, acquisition of motility properties and passage through thin-walled venules and lymphatic channels, both of which offer little resistance to tumor cell invasion; once the cancer cell(s) has detached from the primary tumor, it may intravasate into blood or lymphatic system, both routes of dissemination leading to venous circulation; survival pathways are activated and interactions with the vascular microenvironment are established during tumor cell trafficking through the bloodstream, until tumor cells arrest in a capillary bed and extravasate; after homing to distant sites, successful colonization depends on interaction with the local microenvironment, defense against host immune response and the formation of a pre-metastatic niche by bone marrow-derived hematopoietic progenitors that attract tumor cells and support metastatic outgrowth; cells at secondary lesions can also re-disseminate to re-seed the site(s) of origin or form other metastases (reviewed in Steeg, 2006; Nguyen et al., 2009; Talmadge and Fidler, 2010). Each step can be rate limiting (Figure 1.3.1), as a failure or an impairment of any of these steps can stop the entire process (Fidler, 1970; Poste and Fidler, 1980).

In 1889, Stephen Paget proposed that metastasis depends on the cross-talk between selected cancer cells, defined as “seeds”, and the microenvironment of host tissues, which

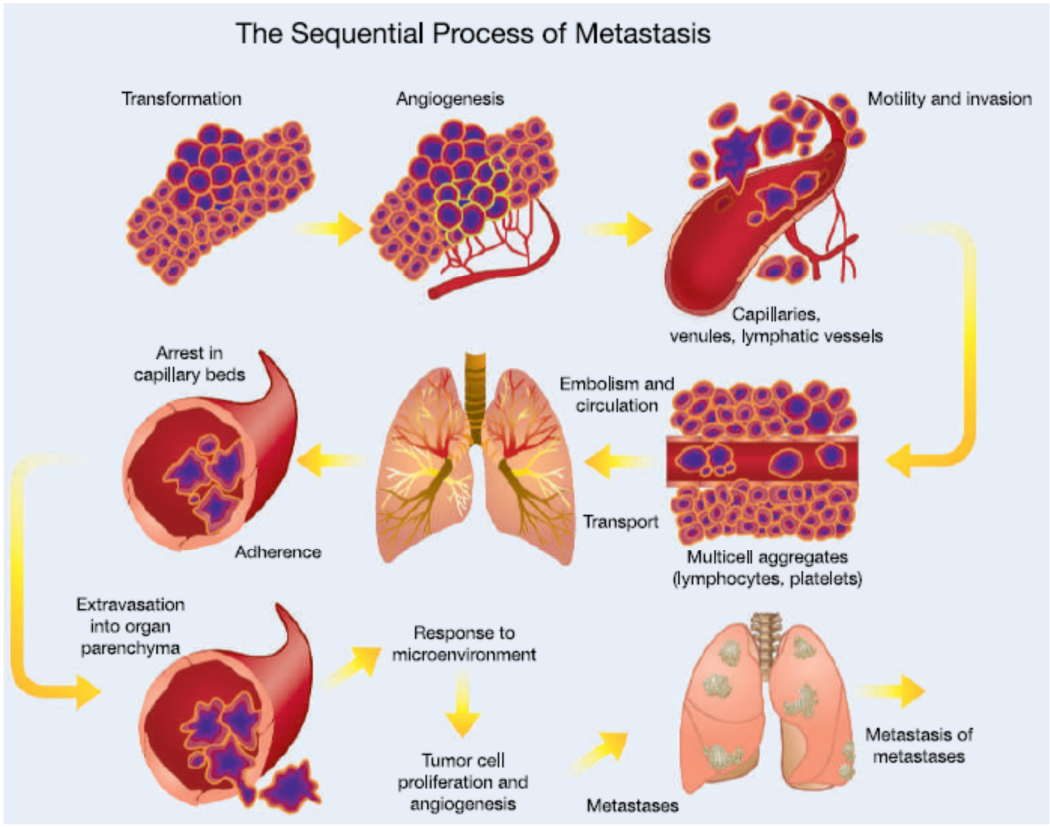


Figure 1.3.1. The metastatic cascade. The process of cancer metastasis consists of sequential and interlinked events, each representing a potentially rate limiting step, such that the formation of clinically relevant metastases depends on the completion of the entire process. *Talmadge and Fidler, 2010*

represents the “soil” (Paget, 1889), on the basis of a series of observations regarding the blood supply and the frequency of metastasis in specific organs, thus contradicting the prevailing theory of Virchow, according to which tumor dissemination is determined by mechanical factors (Virchow, 1989). Forty years later, in 1928, Ewing challenged the “seed and soil” hypothesis (Ewing, 1928). He proposed that mechanical forces and circulatory patterns between the primary tumor and the secondary site accounted for organ specificity. The seminal studies by Fidler and coworkers (Fidler and Kripke, 1977; Hart and Fidler, 1980) conclusively showed that, although tumor cells traffic through the vasculature of all organs, metastases selectively develop in congenial organs. The “seed and soil” hypothesis represents the basis for current research as it is now widely accepted that the outcome of the metastatic process, and consequently of therapeutic protocols, is dependent on both the

intrinsic properties of tumor cells and their interaction with host factors (Quail and Joyce, 2013).

Studies focused on metastasis biology indicated that *in vivo* analyses are required to describe the biological phenomenon at mechanistic level and that in the absence of animal models the probability for misinterpretation is high. On the other hand, clinical studies provide high fidelity models of the timing of dissemination and allow defining intra- and inter-lesion heterogeneity. For instance, studies based on *in vitro* and *in vivo* selection of cells with enhanced metastatic capability demonstrated that cell subpopulations with heterogeneous metastatic potential exist within the same tumor (reviewed in Fidler, 2003). The biological heterogeneity of cells that survive in a foreign microenvironment suggested that the metastatic process is selective and that phenotypic diversification occurs during tumor growth. The British pathologist Leslie Foulds was the first to interpret the heterogeneity of cancer cell populations in a dynamic perspective, describing cancer progression as a stepwise acquisition of permanent and irreversible changes (reviewed in Klein, 1998). This view of cancer was definitively explicated by Peter Nowell, who attributed tumor evolution to acquired genetic variations, suggesting the first explicit Darwinian model of tumor progression (Nowell, 1976). According to the linear progression model, tumor cell phenotypic heterogeneity represents both the effect of cancer cell genetic instability, characterized by multiple rounds of genomic alterations, and of selective pressures exerted from the microenvironment, which determines selective growth of cell clones endowed with enhanced competitive fitness. Accordingly, metastases appear at the latest stage of tumor progression, as a result of the temporal evolution and acquisition of specific molecular traits (Figure 1.3.2). However, this model did not explain the first observations that metastasis may initiate long before the first symptoms appeared or the primary tumor was diagnosed (reviewed in Klein, 2009). On the basis of these considerations, a parallel progression model has been proposed, which places dissemination among the events that occur at the earliest stages of disease by the accumulation of genetic and epigenetic alterations and adaptation to foreign microenvironments in the primary tumor in a parallel and independent manner compared to

metastases (Klein, 2009). Therefore, marked genetic divergence is expected between primary and secondary lesions, including disseminated tumor cells (either circulating in blood or hidden in the bone marrow).

The application of the one or the other model has important implications for clinical studies. In the case of parallel progression, the genetic divergence between primary and metastatic tumors makes the primary tumor unsuitable for diagnostic purposes and for the design of effective therapies. Therefore, the analysis of disseminated tumor cells can provide additional and more useful information rather than studying the primary lesions only. However, even when considering the stepwise progression model, diagnostic and treatment decisions that are based on the analysis of the whole primary tumor and do not take into account the intra-tumor heterogeneity may be inaccurate, since a minor subclone might have determined the formation of metastases.

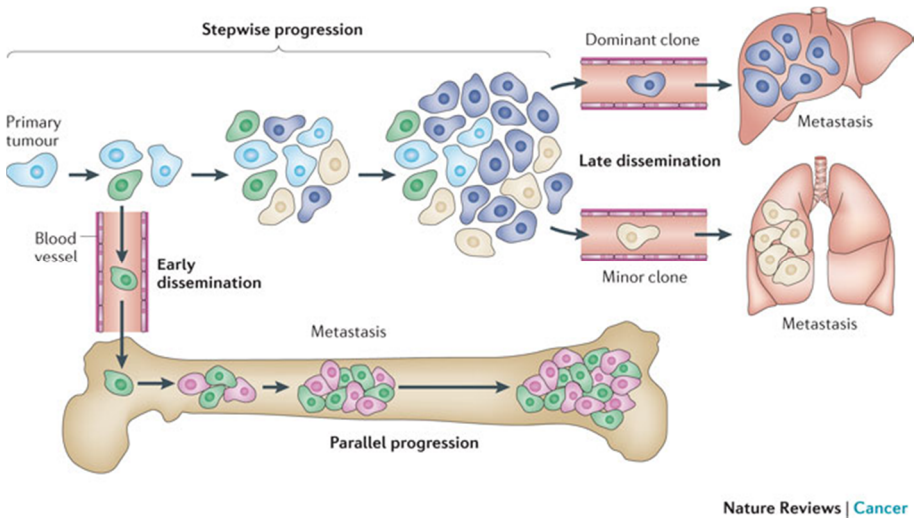


Figure 1.3.2. The linear and parallel progression models. Two hypothetical models of tumor progression during metastasis are stepwise progression, in which metastases are seeded at the latest, clinically diagnosable stage of tumor progression, and parallel progression, in which tumor dissemination occurs at earlier stages of tumor progression. *Marusyk et al., 2012*

Experimental and clinical data are in favor of the early dissemination and parallel progression of breast cancer (Hüsemann et al., 2008; Schardt, 2005), although different behavior of tumor cells regarding the timing of dissemination should be expected due to the

heterogeneity of the disease and the tendency to colonize organs with different characteristics (Klein, 2009).

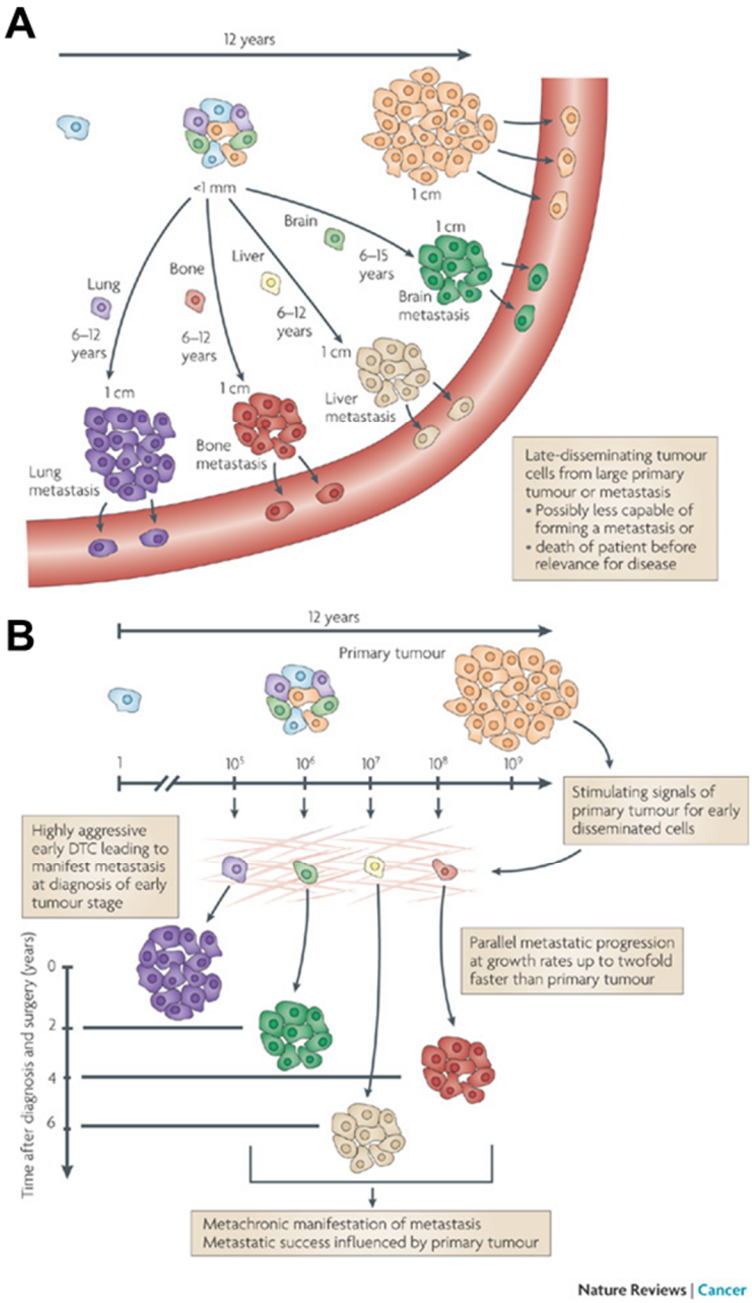


Figure 1.3.3. The parallel progression model in breast cancer. At early stage, (**A**) cancer cells disseminate from tumors with small volumes and give rise to metastases within 6-15 years according to the features of the distant organ. (**B**) Multiple waves of dissemination may occur before diagnosis, under the stimulation of factors secreted by the primary tumor, which can influence the timing and aggressiveness of metastases. *Klein, 2009*

According to the early dissemination and parallel progression model, dissemination starts when tumor diameter is 1–4 mm. Metastases in different organs are seeded in parallel

and develop within 6 to 12 years, with the exception of brain metastases, which can establish later due to organ specific barriers and microenvironment. Late-disseminating tumor cells from primary or secondary tumors with larger volumes are possibly less capable of forming metastases or contributing to poor outcome as death might occur before they can be detectable. Moreover, tumor cells may disseminate in several waves before diagnosis and may progress in parallel at different rates in different organs. Behind cancer cell molecular traits and behavior, factors secreted by the primary tumor may stimulate colonization and account for the relationship of tumor size and probability of metastatic outgrowth (Figure 1.3.3).

1.3.2. Genetic determinants of breast cancer metastases

A salient feature of metastasis is the ability of different tumor types to colonize the same or different organ. However, it remains unclear to what extent these genes are used by different tumor types that show the same pattern of metastasis.

Breast cancer is a heterogeneous disease and also different metastatic behavior was observed according to the molecular subtype. Hormone receptor-positive tumors have a tendency to develop bone metastases and they have better survival outcomes compared to hormone receptor-negative tumors with a tendency of developing visceral metastasis; luminal/HER2-positive and HER2-enriched tumors are associated with a significantly higher rate of brain, liver, and lung metastases; basal-like tumors have a higher rate of brain, lung, and distant nodal metastases but a significantly lower rate of liver and bone metastases; TN non-basal tumors demonstrate a similar pattern but were not associated with fewer liver metastases compared to TN non-basal tumors (Kennecke et al., 2010).

Genetic determinants associated to the different stages of BC metastatic process were identified. Numerous studies revealed that TGF-β signaling may have a multifunctional nature, as clinical and experimental results showed that it may exert a pro-metastatic or a tumor-suppressive role within the same tumor, depending on the cellular context and the

stage of disease (Padua and Massagué, 2009; Massagué, 2012; Kang et al., 2005). In breast cancer TGF-β was proven to mediate the induction of Angptl4 in cancer cells that are going to enter the circulation, with subsequent enhanced retention of these cells in the lungs, but not in the bone; in fact, tumor cell-derived Angptl4, in turns, disrupts vascular endothelial cell-cell junctions, increases the permeability of lung capillaries, and facilitates the trans-endothelial passage of tumor cells (Padua et al., 2008). Metastatic BCs may also evade the growth-inhibitory action of TGF-β by defects in genes that play a key role in the coordination of TGF-β cytostatic response, as the accumulation of C/EBPbeta inhibitory isoform LIP (Gomis et al., 2006).

Active cross-talk between breast cancer cells and bone has been largely documented, and studies in this field allowed defining a vicious cycle which regulates the formation of osteolytic breast cancer metastases to bone (Roodman, 2004; Kozlow and Guise, 2005). Factors, such as MMPs, chemokine receptor 4 (CXCR4), vascular endothelial growth factor (VEGF), and connective tissue growth factor (CTGF), target metastatic tumor cells to bone and facilitate survival within the bone microenvironment. Physical factors within the bone microenvironment, including hypoxia, acidic pH, and extracellular Ca^{2+}, and bone-derived growth factors, such as TGF-β and IGFs, activate tumor expression of osteoblast-stimulatory factors, including VEGF, platelet-derived growth factor (PDGF), and endothelin 1. Osteoclast-stimulatory factors, including PTHrP, TGF-β, and IL-11, can also be increased. These factors stimulate bone cells, which in turn release factors that promote tumor growth in bone. The angiogenesis factor fibroblast growth factor-5, the connective tissue-derived growth factor, the activator of osteoclast differentiation interleukin-11, the matrix metalloproteinase/collagenase MMP1, follistatin, and the metalloproteinase-disintegrin family member ADAMTS1 were also listed among genes which drive tumor cells toward bone colonization (Kang et al., 2003).

A gene signature that regulates breast cancer cell tropism to lung was also identified. ID1, VCAM1 or IL13Rα2 knock-down was proven to decrease the lung metastatic activity of breast cancer cells (Minn et al., 2005). The establishment of distant metastases depends on

the capability of small numbers of cancer cells to regenerate a tumor after entering a target tissue. A role for the transcriptional inhibitors of differentiation Id1 and Id3 as selective mediators of lung metastatic colonization in the TNBC was also reported (Gupta et al., 2007). Moreover, a gain-of-function cDNA screen revealed that Coco, a secreted antagonist of TGF-β ligands, induces dormant breast cancer cells to undergo reactivation in the lung by blocking lung-derived BMP ligands (Gao et al., 2012). In order to decipher the mechanism that regulates tumor cell dormancy and re-awakening, Malladi and co-workers used another approach, based on the isolation of disseminated cancer cells able to survive as latent entities, called latency competent cancer (LCC) cells (Malladi et al., 2016). They demonstrated that LCC cells show stem-cell-like characteristics and express SOX2 and SOX9 transcription factors, which are essential for their survival in host organs under immune surveillance and for metastatic outgrowth under permissive conditions. By expressing a Sox-dependent stem-like state and actively silencing WNT signaling, LCC cells can enter quiescence and evade NK-mediated innate immunity to remain latent for extended periods.

The blood-brain barrier (BBB) exerts a dual role in brain metastasis formation: it forms a tight barrier protecting the central nervous system from entering cancer cells, but it is also actively involved in protecting metastatic cells during extravasation and proliferation in the brain (Wilhelm et al., 2013). Gene expression analysis of cells that preferentially infiltrate the brain isolated from patients with advanced breast cancer, coupled with functional analysis, identified the cyclooxygenase COX2 (also known as PTGS2), the EGFR-ligand HBEGF, and the alpha2,6-sialyltransferase ST6GALNAC5 as mediators of cancer cell passage through the blood–brain barrier (Bos et al., 2009). Once passed through the BBB, tumor cells necessitate the establishment of interactions with the stroma microenvironment in order to fuel their metastatic ability, as reported in a recent paper on the role of protocadherin 7 (PCDH7) (Chen et al., 2016). PCDH7, which is expressed by breast and lung cancer cells, promotes the assembly of carcinoma-astrocyte gap junctions; once engaged with the astrocyte gap-junctional network, brain metastatic cancer cells use these channels to transfer

the second messenger cGAMP to astrocytes, activating the STING pathway and the production of inflammatory cytokines such as interferon-α (IFNα) and tumor necrosis factor (TNF); acting as paracrine signals, these factors activate the STAT1 and NF-κB pathways in brain metastatic cells, thereby supporting tumor growth and chemoresistance.

Information on the mechanisms that regulate the late phase of BC metastasis, including colonization of distant organs and metastasis outgrowth, represent a large part of literature; however, data on the mechanisms regulating the early events of dissemination and survival of tumor cells in blood are still scarce.

1.3.3. Modeling breast cancer metastases in the mouse

Studies carried out in recent years have brought impressive advancements in BC metastasis research by the use of a large number of experimental models, including drosophila, zebrafish, mice, rats and more rarely, rabbits, companion pets and monkeys (reviewed in Saxena and Christofori, 2013). Metastasis is a biological process that requires the completion of a cascade of events. The main hallmark properties of a metastatic cell are its migratory and invasive abilities (Yang and Weinberg, 2008; Friedl and Wolf, 2010). *In vitro* assays, such as scratch wound, Boyden chamber migration, transendothelial migration, and Matrigel invasion (Stoker and Perryman, 1985; Boyden, 1962; Reiske et al., 1999; Shaw, 2005; McClatchey, 1999; Lee et al., 2007), have been employed to investigate on the molecular mechanisms that orchestrate these processes. Also experimental systems that more reliably represent the tumor microenvironment, as the implantation of tumor cells into the chorioallantoic membrane of the chick embryo, allowed the live monitoring of tumor cell invasion, as also angiogenesis and interaction between cancer cells and blood vessels (Sherman et al., 1998).

While *in vitro* model systems are instrumental for the analysis of single functional aspects of cancer cells, the information gained on the complexity of the metastatic process is limited as it is free from the biological context offered by an entire organism. A major

contributor to deciphering of the multistage nature of metastasis comes from transgenic and transplantation laboratory mouse models, which have served as important preclinical tools for both mechanistic and pharmacological studies (Bos et al., 2010). Cancer models driven by the introduction of oncogenic mutations in a tissue-specific manner can faithfully recapitulate important aspects of tumor initiation, local progression, and response to therapy (Bos et al., 2010). In BC models as those MMTV-driven HER2-transgenic or MMTV-driven PyMT-transgenic (which resemble the luminal B molecular subtype, Herschkowitz et al., 2007), cancer develops with high penetrance in a stepwise manner, enabling the study of tumor initiation and of the early steps of metastatic dissemination (Lin et al., 2003; Ursini-Siegel et al., 2007). The MMTV-PyMT model shares many aspects of human BC progression, and multistage progression from hyperplasia to multifocal mammary adenocarcinomas followed by the development of metastatic lesions in lymph nodes and lung with high penetrance and short latency (Guy et al., 1992; Maglione et al., 2001). The MMTV-Her2/*neu* mice develop multifocal adenocarcinomas with lung metastases at about 15 weeks after pregnancy (Muller et al., 1988). In a seminal paper, Hüsemann and colleagues found that in MMTV-Her2/*neu* mice cancer cells leave the primary tumor site and reach bone marrow and lung before cancer become morphologically detectable (Hüsemann et al., 2008).

Another exemplary model is represented by tail vein injection of mammary epithelial cells genetically modified for several oncogenes under control of a doxocyclin-inducible promoter, which showed that few untransformed epithelial cells may colonize the lung once they have entered the bloodstream and are capable of surviving and form colonies at lung upon oncogene activation (Podsypanina et al., 2008). However, metastasis in these models is often restricted to lymph nodes and lungs.

Syngeneic and xenograft models in which mouse or human cancer cells are introduced into immunocompatible or immunocompromised mice, respectively, provide at present two methods of choice to experimentally induce metastatic dissemination and colonization of relevant organs.

Syngeneic BC cell lines were derived from a single spontaneous tumor arising in BALB/cfC_3H mouse (Dexter et al., 1978; Miller et al., 1983), and murine BCCLs with different metastatic potential were then obtained: the 66cl4 (Miller et al., 1983), 67NR and 168FARN (reviewed in Heppner et al., 2000) variants were derived from 66, 67 and 168 lines, respectively, and showed low metastatic potential (Aslakson and Miller, 1992), whereas 4T07 and 4T1 variants were derived from the 410.4 line and, notoriously, metastasize to lung, liver and lymph nodes (Dexter et al., 1978; Aslakson and Miller, 1992; Heppner et al., 2000), when injected in Balb/c mice. Clone 4T1.2 (and others), showing tropism also to bone, was derived from the 4T1 cell line, (Lelekakis et al., 1999; Eckhardt et al., 2005). The various sublines of 4T1 have also been used to generate distinct gene expression signatures for each stage of tumor progression, and comparison of their profiles has led to the identification of the EMT-inducing transcription factor Twist, which was also shown to be important for BC metastasis (Yang et al., 2004), as also the BMP inhibitor Coco, able to re-awake 4T07 dormant cells at lung in Balb/c mice (Gao et al., 2012). Importantly, transplantation of cancer cells in mice with identical genetic background bypasses the immunologic host-versus-graft rejection and concomitantly allows investigating the contribution of an intact immune system and other players of the tumor microenvironment to malignant tumor progression (Khanna and Hunter, 2005).

Metastatic xenograft mouse models have been established also for BC studies, in particular with the MDA-MB-231 cell line and its derivatives, providing a suitable system to investigate the late steps of the metastatic cascade (Nguyen et al., 2009). In experimental metastasis assays, the tumor cells are injected directly into the systemic circulation. The development of metastasis in these models is rapid and is largely influenced by the site of injection and the inherent tropism of the tumor cells, especially after *in vivo* selection of organ-specific metastatic variants after several rounds of transplantation (Nguyen et al., 2009; Bos et al., 2010). The various routes include lateral tail vein, intra-portal, intra-splenic, intra-carotid, intra-peritoneal and intra-cardiac injections (Khanna and Hunter, 2005). Injection into the tail vein is the most commonly used assay and primarily leads to formation

of metastatic nodules in the lung (Minn et al., 2005), whereas intra-peritoneal injection usually provokes local invasion (Khanna and Hunter, 2005). Intra-portal vein and intra-splenic injections usually result in liver metastasis, intra-carotid injection leads to brain (Bos et al. 2009), and left ventricle injection favors multiorgan metastases (Harms and Welch, 2003; Khanna and Hunter, 2005), including bones (Yin et al., 1999; Kang et al., 2003).

In spontaneous metastasis assays, cancer tissue or tumor cells are primarily implanted into the organ from which the cancer cells have been originally derived (orthotopic transplantation) or into a tissue of high vascularization and convenient anatomical location that does not represent the organ of origin (ectopic transplantation), such as the skin or the subrenal capsule (Khanna and Hunter, 2005). Ectopic transplantation models mostly fail to mimic the appropriate microenvironment of the primary tumor and the corresponding metastatic dissemination to the relevant organs. Hence, orthotopic transplantation models have been developed that more closely mimic the human situation, including tumor histology, vascularity, tumor-stroma interactions, gene expression profiles, metastases development, and also chemotherapy responsiveness (Céspedes et al., 2006). With spontaneous models, most stages of the metastatic cascade can be observed and investigated (Saxena and Christofori, 2013).

Despite the recognized importance of the previously described models, BC modeling remains a challenge due to the intrinsic heterogeneity of this neoplasm. In order to successfully identify new therapies, it is fundamental that preclinical models could recapitulate the complexity observed in the clinical context (Landis et al., 2013; Whittle et al., 2015). Improvements in engraftment protocols and utilization of more severely immunocompromised mouse models set the stages for advancements in generating patient-derived xenograft (PDX) models of BC. Marangoni and colleagues (Marangoni et al., 2007) achieved a 15% and 24% initial engraftment rate from primary tumors and metastatic sites, respectively. These Swiss nude mouse models produced a 12.5% (25/200) stable take rate and ten models with lung metastases. In comparing xenograft response to patient

response to treatment, a 5/7 concordance was observed, supporting the utility of these models for evaluating therapeutics. DeRose and colleagues obtained metastases after engraftment of biospecimens of primary tumors, pleural effusates and ascites, with frequencies from 38% to 100% in sites corresponding to those diagnosed in patients (DeRose et al., 2011). Zhang and colleagues established a large cohort of 35 stable cell lines representing 27 patients, with 12 lines (48%) developing metastatic lesions in the lung (Zhang et al., 2014). Regardless of the challenges in establishing PDX, these models generally retained the pathological characteristics and biomarker status of the original patient tumor and were proven to be stable across multiple transplant generations. Several groups have documented retention of histopathological characteristics and biomarker status by IHC, whereas molecular analysis, although showing shared alterations between primary and xenograft tumors, revealed more pronounced mutational status or aggressiveness tumor characteristics in PDXs (Zhang et al., 2014), as also observed by Ding and colleagues by comparing deep sequencing results of peripheral blood, the primary tumor, a brain metastasis and a xenograft derived from the primary basal-like TN breast tumor (Ding et al., 2010). More recently, the use of CTCs isolated from small cell lung cancer and prostate cancer patients has been also adopted as reliable model to study tumor aggressiveness and chemosensitivity, and to provide tractable systems for therapy testing (Hodgkinson et al., 2014; Williams et al., 2015; Alix-Panabières et al., 2016). A major challenge to the establishment of PDXs from BCs is the low take rate of luminal tumors compared to TNBC (reviewed in Landis et al., 2013). In addition, xenograft models in general do not allow studies on the role of the microenvironment and the immune system in metastasis and therapy response.

Overall, all of these models provide complementary tools to study BC metastasis biology since no single metastasis model is sufficient to answer all questions. As such, the selection or development of suitable models is necessary.

1.4. TREFOIL FACTOR 3

1.4.1. Structure and physiopathological functions of trefoil factors

The trefoil factor (TFF) peptides consist of a family of three highly conserved, thermostable and protease-resistant secreted polypeptides (Thim and May, 2005): human TFF1 (formerly pS2), TFF2 (formerly spasmolytic polypeptide/SP), TFF3 (previously called intestinal trefoil factor/ITF or hP1.B). TFF1 was the first member of the family to be discovered via cDNA cloning of an estrogen-responsive gene from a BC cell line (Masiakowski et al., 1982); human TFF2 was characterized few years later (Tomasetto et al., 1990); human TFF3 from rat intestine was the third mammalian TFF-peptide described (Suemori et al., 1991) and its human homolog was reported in 1993 (Hauser et al., 1993; Podolsky et al., 1993).

Figure 1.4.1. Trefoil domain 3D structure. Picture represents the predicted 3D structure of TFF3 (http://prosite.expasy.org/PS51448); trefoil domain is highlighted in green.

The protein family was named for the disulphide bond configuration of the P domain, or trefoil domain, which forms a three-leaved structure similar to a trefoil or clover leaf

(Figure 1.4.1), and the official nomenclature was assigned upon an agreement at a *Conférence Philippe Laudat* (Wright et al., 1997).

The trefoil domain was originally defined as a conserved sequence of 38 or 39 amino acidic residues, containing 6 cysteine residues forming disulphide bonds in a 1-5, 2-4 and 3-6 configuration, which are responsible for the remarkable protease resistance of TFF peptides. Later, examination of the available three-dimensional structure and comparison of the conservation between mammalian TFFs revealed that the conserved P domain extends beyond the 38-39 amino acidic residues originally suggested, and protein structure data revealed that trefoil domain has a sequence of 42-43 amino-acidic residues. The analysis of the tertiary structure also showed that the 3 loops do not lie on a single plane but are stacked parallel to each other (reviewed in Thim and May, 2005).

Genes encoding for TFF peptides are located as a head to tail oriented cluster in a 55 kb region on the reverse strand of chromosome 21q22.3. The 5'-flanking regions of these genes have similar regulatory sequences, suggesting that the three genes are transcribed and regulated in a coordinated fashion (reviewed in Regalo et al., 2005). *TFF1* gene has 1 transcript with 3 exons encoding for a 84 amino-acidic residues polypeptide and the mature peptide consists of 60 residues containing a single trefoil domain. *TFF2* gene has 3 transcripts, 1 of them encoding for a 129 amino-acidic residues polypeptide, while the mature sequence consists of 106 residues containing 2 trefoil domains, encoded by 2 separate exons. *TFF3* gene has 4 transcripts, 3 of them encoding for a 94, a 125 and a 74 amino-acidic residues polypeptide, respectively, but the 94-residues peptide, consisting of 59 amino-acidic residues and containing a single trefoil domain, seems to be the most represented and supported variant, according to Transcript Support Level and APPRIS computational methods (www.ensembl.org). Residues from 1-24, 1-23 and 1-21, respectively for TFF1, TFF2 and TFF3, encodes for signal peptide (www.uniprot.org).

Both TFF1 and TFF3 form homodimers via disulphide bond with a seventh cysteine residue and can form disulphide bonds with other proteins (Westley et al., 2005; Chadwick

et al., 1997, May et al., 2003). TFF2 has two extra-trefoil domain cysteine residues that form an intramolecular disulphide bond (Thim, 1989).

According to overall microarray, RNAseq and SAGE gene expression profile (GEP) data (www.genecards.org), *TFF1* shows high expression in the stomach and colon and moderate expression in the prostate, small intestine, pancreas, thyroid, salivary gland and breast; *TFF2* shows high expression in the stomach and moderate expression in the thyroid, placenta, salivary gland and pancreas. Overall, both peptides are expressed at low level in other visceral organs, in the nervous system, muscle and hematopoietic cells. *TFF3* is expressed at higher level in the small intestine, colon, thyroid and pancreas compared to the mammary gland, ovary, uterus, prostate, liver, kidney, bladder, testis, salivary gland, adipocytes and bone marrow, where moderate signals were observed, whereas it can be detected at low levels in the nervous system and muscle. No peptide expression was found in white blood cells and serum, whereas TFFs are all detectable in urine and pancreatic juice (www.genecards.org). TFF3 expression body map is reported in Figure 1.4.2.

TFF peptides have a pivotal role in maintaining the surface integrity of a number of mucin-producing epithelia, with TFF1 and TFF2 principally expressed in gastric mucosa and TFF3 manly secreted by goblet cells of intestine and lung (reviewed in Taupin and Podolsky 2003). They can be co-secreted with mucins into the lumen to form a gel-like mucus layer and their association with mucins exerts a protective effect on the mucosa, serving as a barrier from mechanical stress, noxious agents and pathogens (Hoffmann 2009; Kjellev 2009). Consistently, TFF3 binding on epithelial cell surface is regulated by mucins, TFF2 has been shown to catalyze the formation of stable mucin complexes and TFF1 interacts with MUC5AC through binding to the von Willebrand C cysteine-rich domains VWFC1 and VWFC2. However, TFF peptides can also act at intracellular level, where they are believed to contribute to protein folding. Indeed, the lack of TFF1 leads to the accumulation of misfolded proteins in the endoplasmic reticulum.

TFF peptides are rapidly up-regulated and secreted in response to gastrointestinal injuries and act as motogens in autocrine and paracrine fashion to facilitate cell migration into

the lesion, thus forming a protective barrier in a process known as restitution (Poulsom et al., 1996). In addition, they are potent inhibitors of apoptosis and prevent *anoikis* during the cell migration process (Taupin et al., 2000; Figure 1.4.3).

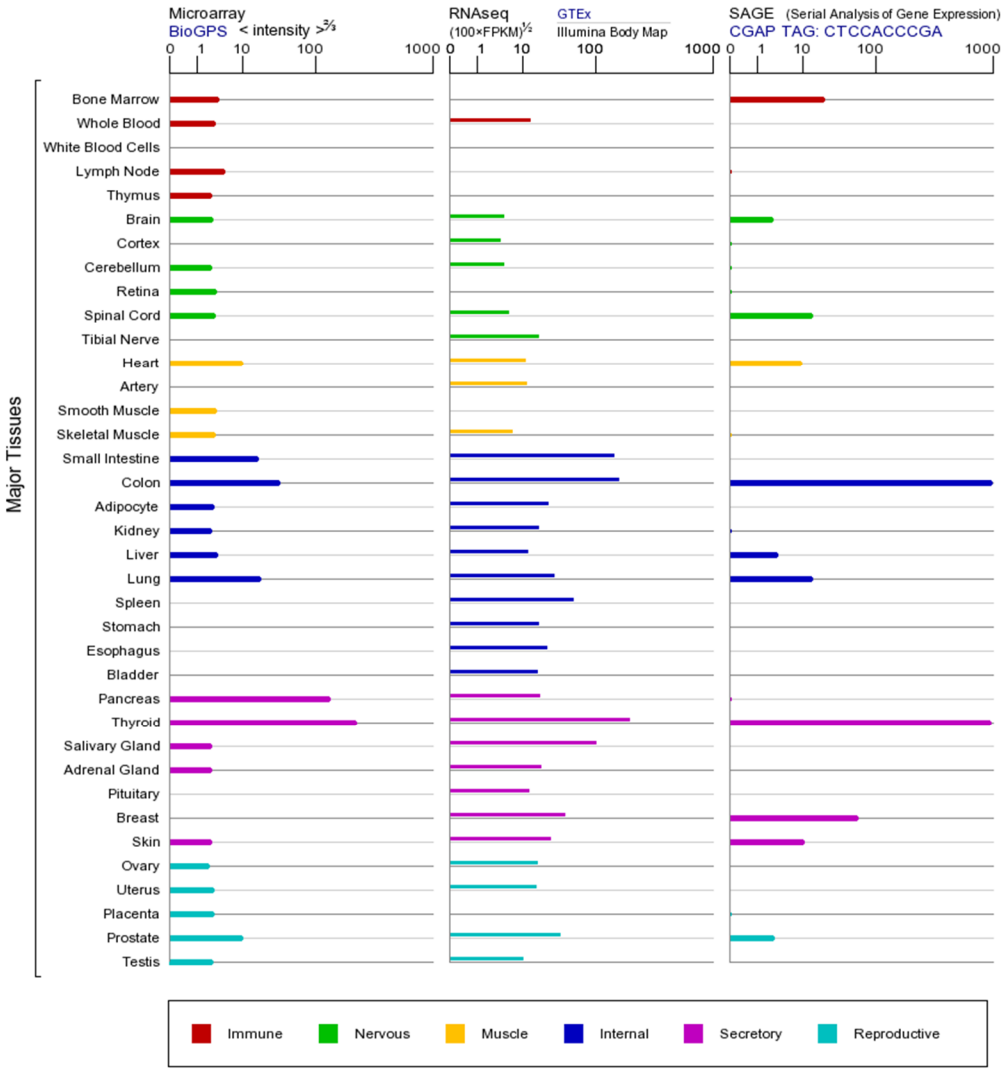


Figure 1.4.2. TFF3 expression body map. Map reports expression scores based on data from GTEx, BioGPS, Illumina Human BodyMap, and SAGE (www.genecards.org).

Aberrant expression of TFF-peptides was widely observed during various chronic inflammatory diseases of epithelial tissues, as for example in ileal Crohn's disease, in colon mucosa of patients with ulcerative colitis and in gallbladder during acute cholecystis (reviewed in Poulsom and Wright, 1993). Moreover, all TFF peptides are typically secreted by a specific gland-like structure termed the ulcer-associated cell line (UACL) (Poulsom and

Wright, 1993). This mucin-secreting glandular structure appears during a variety of chronic inflammatory conditions and contains a clearly defined proliferative zone which develops from stem cells, most commonly in the small intestine, and is thought to represent a natural repair system which is activated after mucosal damage, particularly in Crohn's disease and duodenal ulcer disease.

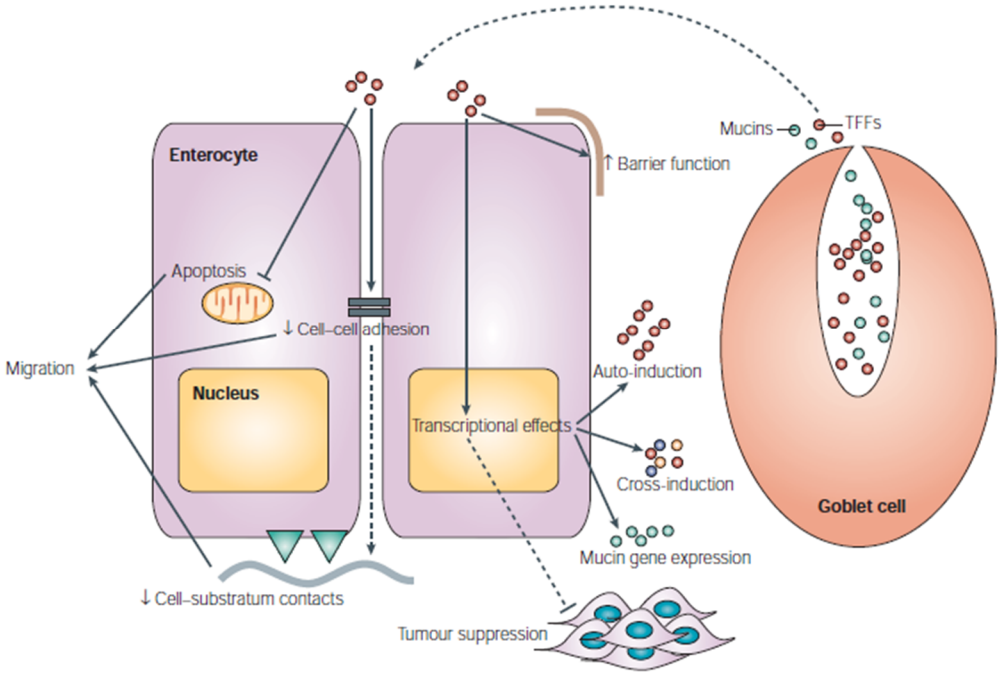


Figure 1.4.3. Established and putative functions of trefoil proteins. Established functions are indicated by solid arrows, whereas putative functions are denoted by dashed arrows. Mucous-cell (for example, goblet cell) secretion of trefoil factors (TFFs; red circles) or mucins (green circles) is concomitant. Secreted TFF acts on adjacent mucosal cell populations (for example, enterocytes) either at extracellular or intracellular level. Cell migration is the result of integrated disruption of cell-cell and cell–substratum adhesion and prevention of *anoikis*. TFF response elements in TFF gene promoters allow increases in TFF expression through auto-induction (red circles) and cross induction of other TFFs (blue and yellow circles), in addition to mucin expression and possibly tumor suppression. *Taupin and Podolsky, 2003*

1.4.2. Role of trefoil factors in cancer

In addition to the protective and restorative effects of TFFs in the gastrointestinal tract, recent evidence has indicated that members of this family also possess pivotal role in the development and progression of human cancers. Aberrant expression of TFFs has been

reported for a variety of solid tumors, including breast, prostate, colon, lung and gastric cancers (reviewed in Perry et al., 2008).

In the current literature a role as tumor suppressor genes and potential tumor progression factors is reported for TFF peptides. The reason for this apparent contradiction might be a consequence of the complex relationships between the inflammatory environment frequently observed in the digestive tract during neoplastic transformation (included generation of reactive oxygen species, associated genotoxic damage, cellular stress and mitogenic signals) and oncogenic activation of key genes involved in cellular transformation and tumor progression (Emami et al., 2004).

In support of a tumor suppressor role, both TFF1 and TFF3 reduce the growth of the human colon cancer cell lines LoVo and SW837 *in vitro* and *in vivo* (Calnan et al., 1999; Uchino et al., 2000). More recently it was demonstrated that deficiency in TFF1 increases tumorigenicity of human breast cancer cells and mammary tumor development in TFF1-knock-out mice (Buache et al., 2011).

However, as also summarized in Table 2, TFF peptides were proven to promote cell migration and resistance to *anoikis* and antiapoptotic role was also observed after serum starvation in *in vitro* assays. They also act as motogens both in normal and malignant states and promote invasiveness and angiogenesis (references are listed in Table 2). Moreover, TFF3 enhances cell proliferation and survival and reduces sensitivity to ionizing radiations in prostate carcinoma cells (Perera et al., 2015). Further functional *in vitro* studies revealed that forced expression of TFF3 increased cell proliferation and survival, enhanced anchorage-independent growth, promoted migration and invasion (Kannan et al., 2010; Pandey et al., 2014) and stimulated endothelial cells adhesion, transmigration through and endothelial barrier and *de novo* angiogenesis (Pandey et al., 2014; Lau et al., 2015). Moreover, forced expression of TFF3 in MCF7 cells stimulated the formation of metastatic nodules at lung when injected in mice, and consistently, siRNA-mediated depletion or neutralization of secreted *TFF3* by antibodies reduced invasive potential, promoted

apoptosis, decreased cell growth *in vitro* and arrested tumor growth in xenograft models (Kannan et al., 2010; Pandey et al., 2014).

Regarding the oncogenic role of TFF3, no evidence exists for the TFF3–mediated acquisition of tumor initiation ability in animal models.

TFF3 expression has also been explored in clinical samples and tested as potential biomarker to predict BC outcome. Immunohystochemistry analysis revealed variable expression of TFF3 in breast tissues. Ahmed *et al.* described (Figure 1.4.4) strong TFF3 immunoreaction in most of the normal breast tissue from premenopausal and perimenopausal women. As in normal tissues, TFF3 was expressed by epithelial - and not myoepithelial - cells of benign breast lesions, and it was especially concentrated in the lumen, where cells appeared with large secretory granules, suggesting active release of TFF3 in the lumen. It was detected in 89% of *in situ* and in 83% of invasive carcinomas, although TFF3 was concentrated at the luminal edge in well-differentiated tumor cells, whereas in poorly differentiated tissues it preferentially accumulated in the stroma. In invasive malignant lesions the immunoreaction was cytoplasmic with a tendency toward perinuclear condensation in some cells, and cells with strong TFF3 expression were scattered diffusely throughout the stroma, whereas in tumors with weaker expression they had a tendency to cluster or aggregate together within the stroma. Within these clusters, malignant cells with higher TFF3 expression were concentrated on the edges of the cell aggregates and in some cases had polarized TFF3 immunoreaction that was more intense toward the stromal edge of the tumor cell. The authors of this work also observed that TFF3 expression was high in well-differentiated tumors and was significantly associated with low histological grade and with ER and PgR expression, which is in line with the induction of *TFF3* mRNA by estrogen in BC cells. Paradoxically, notwithstanding association of TFF3 with features of good prognosis, its expression was also associated with muscle, neural and lymphovascular invasion, and it was an independent predictive marker of lymph node involvement (Ahmed et al., 2012). Consistently with an angiogenic function (Lau et al., 2015), TFF3 expression correlated strongly with microvessel density evaluated with CD31 and

CD34 (Ahmed et al., 2012). In an independent study on two cohorts of BC patients, TFF3 expression was also positively associated with larger tumor size and reduced relapse-free and overall survival (Pandey et al., 2014).

The association between TFFs and ER is not completely clear. *TFF3* - and also *TFF1*, but not *TFF2* - mRNA was detected in breast tumors and estrogen-responsive BCCLs and it was also observed to be induced in an estrogen-dependent and progesterone-independent manner (May and Westley, 1997). Furthermore, *TFF3* was among the top genes whose expression correlated strongly with that of ER in large cohorts of BC patients (Gruvberger et al., 2001). Consistently, forced expression of TFF3 additively increased ER transcriptional activity in the presence of estradiol, promoted estrogen-independent growth and induced resistance to tamoxifen and fulvestrant (Kannan et al., 2010). However, GEP analysis of ER-negative/PgR-negative tumors allowed identifying *TFF3* among a panel of markers characterizing a distinct subset of tumors, showing paradoxical expression of direct ER target genes, as also genes responsive to estrogen or typically expressed in ER-positive BCs (Doane et al., 2006). Interestingly, *TFF1* and *TFF3* were also differentially expressed between BC patients who experienced relapse to bone versus those who experienced relapse in other parts of the body (Smid et al., 2006), and *TFF3* was also detected in tumor cells isolated from the cerebrospinal fluid of metastatic BC patients with leptomeningeal carcinomatosis (Magbanua et al., 2013).

In an attempt to move from the analysis of solid tumor lesions toward the possibility to develop a blood test for detection circulating biomarkers, Bosma *et al.* identified a series of genes abundantly expressed in BC tissues but absent in normal peripheral blood cells and bone marrow, including *TFF3* and *TFF1* (Bosma et al., 2002). Few years later Smirnov and colleagues analyzed the GEP of at least 100 tumor cells isolated from the blood of one metastatic breast, one metastatic colorectal and one metastatic prostate cancer patient and they found that both *TFF1* and *TFF3* were among genes distinguishing between patients and healthy donors (Smirnov et al., 2005).

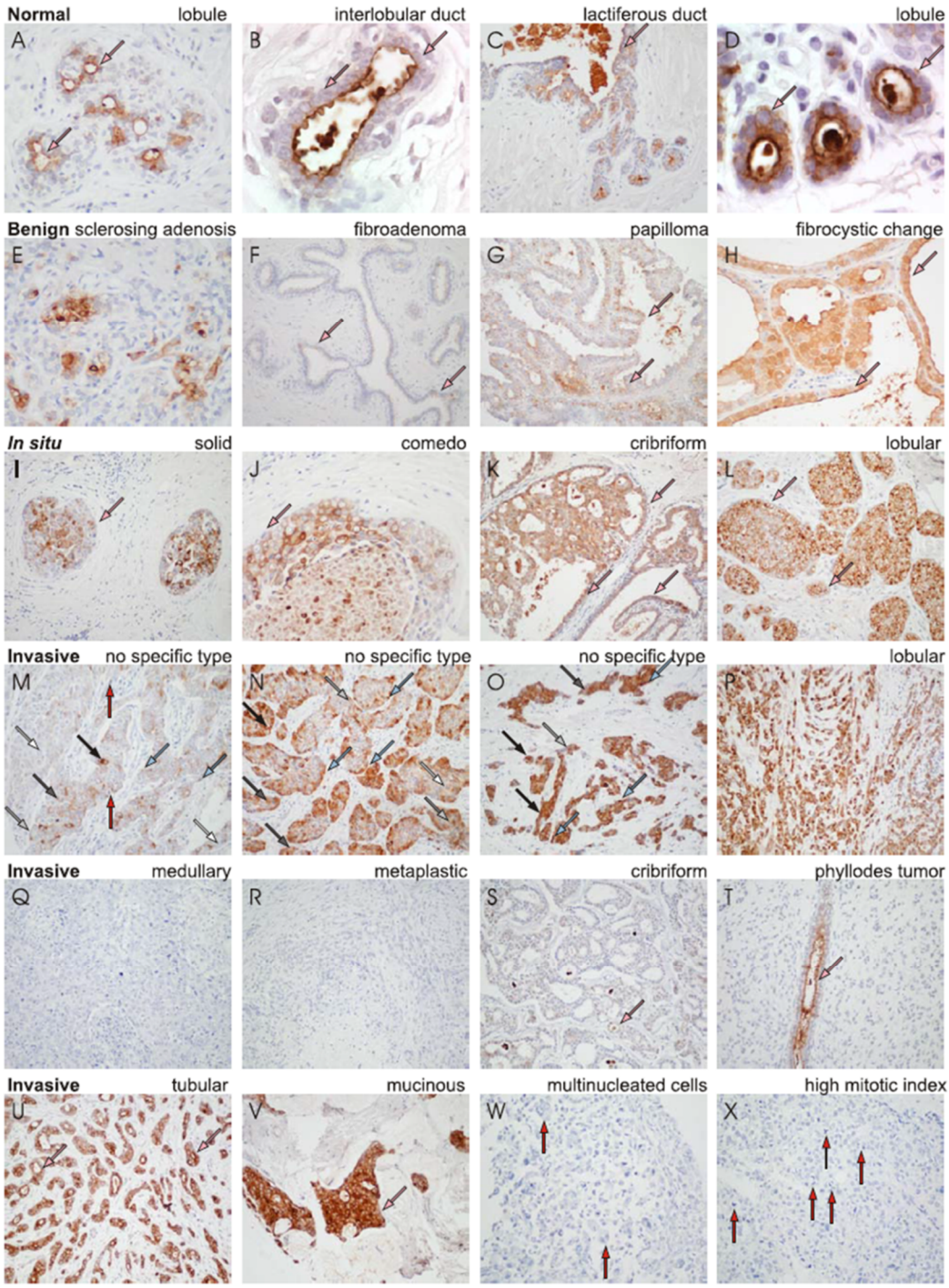


Figure 1.4.4. Expression of TFF3 in normal and diseased human breast. TFF3 protein expression was analyzed in different benign and diseased tissues by Ahmed and co-workers. Arrows indicate malignant epithelial cells with TFF3 expression absent (**M** and **N**; white), weakly positive (**M–O**; pale gray), moderately positive (**M–O**; intense gray), and strongly positive (**M–O**; black). Examples of cells in which TFF3 expression is polarized either away from the lumen (pink arrows) or away from the stroma (blue arrows) are indicated. Multinucleated cells in a carcinoma with prominent cytologic atypia (**W**), and mitotic cells in a carcinoma with mild cytologic atypia (**X**) and in a ductal carcinoma of no specific type (**M**) are indicated (red arrows). *Ahmed et al., 2012*

More recently, a multi-marker panel including both *TFF1* and *TFF3* was developed to identify circulating tumor cells (CTCs) in the blood of metastatic BC patients, revealing that *TFF3* was 10 to 15-fold more expressed in the peripheral blood mononuclear cell fraction isolated from patients compared to healthy controls (Lasa et al., 2013). Interestingly, dynamic changes of CTCs in the epithelial and mesenchymal composition were recently studied in BC patients, showing that tumor cells with epithelial features, which were highly representative of the CTC population in ER-positive/PgR-positive cases and persistent in patients with initial response to therapy, overexpressed *TFF1* and *TFF3* compared to tumor cells with mesenchymal or intermediate features (Yu et al., 2013).

CK+/CD45- CTCs isolated from metastatic BC patients were also characterized by high-resolution copy number profiling: unsupervised clustering of probe level copy number data allowed identification of two CTC-genomic signatures consisting of recurrent copy number gain, either including dormancy-related or tumor aggressiveness-related genes (Kanwar et al., 2015). The two signatures were almost mutually exclusive, except for two copy number aberration regions including among others also *TFF3* gene, thus suggesting that genes located in these regions harbored aberrations early in the hierarchy of cancer cell clones in order to provide CTCs with a survival advantage in the bloodstream and thus a higher metastatic potential (Kanwar et al., 2015).

The role of TFF peptides in cancer is still controversial and probably it cannot be generalized. TFFs function in tumor progression should be investigated according to the type and stage of disease, and the expression of a specific member of the TFFs family is worth being explored both at mRNA and protein level and in different tissues.

1.4.3. Molecular regulators and signaling pathways linked to trefoil factors

Several observations indicate that TFF expression is modulated by many different factors. Among them, *cis*-regulatory elements were identified. Goblet-cell-specific transcription is partly conferred through a nine-base-pair goblet-cell response element

(GCRE) that is present in the proximal *TFF3* promoter region (Ogata et al., 1998). Adjacent positive and negative regulatory elements also contribute to this selective expression (Iwakiri and Podolsky, 2001). Another element, designated the goblet-cell silencer inhibitor (GCSI), is located 2,216 base pairs upstream of the transcriptional start site and interacts with a GCSI-binding protein (GCSI-BP); interaction of this GCSI-BP with the GCSI seems to underlie the induction of *TFF3* expression that is seen after stimulation of undifferentiated cells of the colon cancer cell line HT-29 (Iwakiri and Podolsky, 2001).

A plethora of transcription factors (TFs) and environmental and chemical agents were also proven to activate the *TFF* promoter genes (reviewed in Emami et al., 2004). Regarding *trans*-regulatory elements, analysis of human and mouse *TFF* genes has identified consensus sequences for binding of promiscuous transcription factors. The consensus HNF3/Forkhead-binding motif lies within 100 base pairs of the transcriptional start sites of all three human and mouse TFFs, and HNF3 was shown to bind the *TFF1* promoter and strongly activate *TFF1* transcription in reporter-gene assays. Consensus binding sites for zinc-finger transcription factors have also been characterized in *TFF* promoters, and GATA6, in particular, was shown to activate reporter-gene expression driven by *TFF1* and *TFF2* promoter sequences. Also USF, a member of the basic-helix-loop-helix leucin zipper family of transcription factors, was proved to bind *TFF* promoters in cell lines and to regulate the activation of endogenous *TFF2*. Moreover, promoter analysis of the human *TFF1* gene showed the presence of an estrogen-response element as also enhancer sites responsive to EGF, tissue plasminogen activator and the c-*ras* and c-*jun* oncogene products, whereas *TFF3* expression is induced in an ER-dependent but PgR-independent manner.

Signaling pathways triggered by TFF peptides are still poorly known, also because putative TFF receptors or recognition sites have not been clearly identified yet. TFFs are fundamental players of a physiological process called epithelial cell restitution, thus they act as motogens, promoting cell migration, as pro-angiogenic factors, and prevent anchorage-dependent cell death. Multiple signaling pathways appear to be linked to the biological actions of TFF peptides, including PI3K/AKT pathway, the Rho-ROCK cascade,

COX-2/TXA2-R/Gaq signaling, PLC/PKC, MAP kinases and EGFR signaling. TFF-related signaling pathways were summarized in Table 2 and a model of molecular cascades triggered by the TFF peptides is reported in Figure 1.4.5.

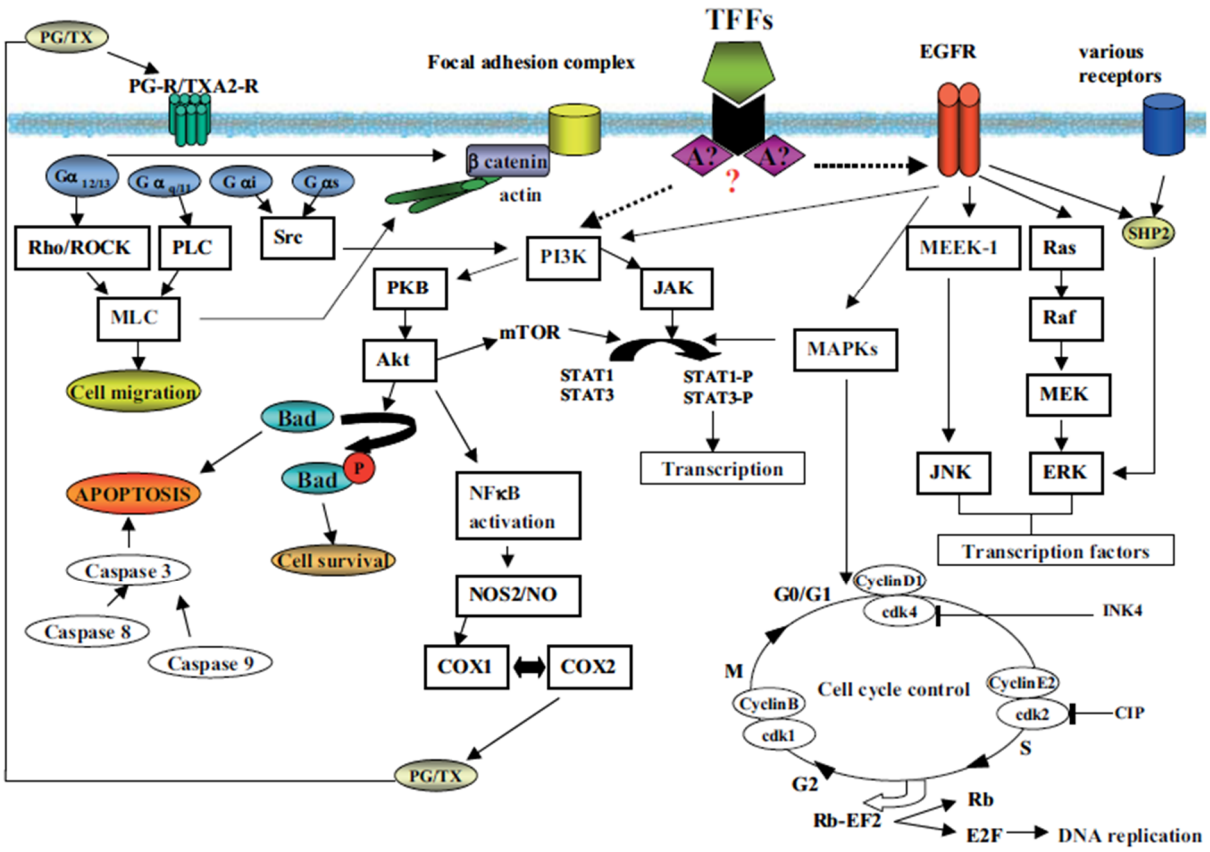


Figure 1.4.5. Proposed model for TFF signal transduction pathways in neoplastic progression and tumor cell invasion. TFF effects are transmitted by unknown adaptor (A?) proteins. TFFs can act through the EGF receptor, which activates several downstream effector pathways such as PI3K/PKB Akt, RAS/Raf/MEK/ERK and MAP kinases MEEK1/JNK. TFF signaling affects apoptosis, cell survival and cell cycle progression through regulation of transcription factors such as NF-κB, STAT and STAT3. TFF signaling involves iNOS (inducible nitric oxide synthase)-stimulated activation of COX1 an COX2, whose prostaglandin products are able to activate PG-R/TXA2-R pathway. Their action is transmitted by heterotrimeric G-protein subunits and activates various cascades such as Rho/RO and PLC, leading to cell migration and invasion, and Src activation that activates NF-κB and PKB/Akt regulated BAD phosphorylation, thus promoting cell survival. TFF1 reduces the activity of caspase-3, -8 and -9, and inhibits apoptosis. TFF1 also delays cell cycle progression by inducing cyclin-dependent kinase inhibitors INK and CIP. TFF peptides activate several signaling pathways leading to cell migration, survival, angiogenesis and, when overexpressed, to cell transformation. *Emami et al., 2004*

Table 2. TFF-related signaling pathways involved in cancer

TFF factor	Biological process			Reference(s)
	SURVIVAL/APOPTOSIS/*ANOIKIS*	**MIGRATION/INVASION**	**ANGIOGENESIS**	
TFF1	• Growth inhibition in colon cancer via MAPK/ERK phosphorylation and increased levels of p16INK4 and p21CIP1/Waf1 (1)	• Cellular invasion dependent on EGFR tyrosine kinase activity in kidney and colonic cancer (4) • Cellular invasion abolished by inhibition of STAT3 signaling in colonic cells (5) • Cellular invasion by induction of COX-1 and COX-2 in colonic cells (8)	• Angiogenesis induction in a COX-2 and EGFR-dependent manner in chick chorioallantoic membrane assay and also via KDR/flk-1 in human umbilical vein endothelial cells (3) • Angiogenesis by induction of COX-1 and COX-2 in colonic cells (8)	(1) Calnan et al., 1999; (3) Rodrigues et al., 2003b; (4) Rodrigues et al., 2003a; (5) Rivat et al., 2005; (8) Rodrigues et al., 2001
TFF2		• Cellular invasion dependent on EGFR tyrosine kinase activity in kidney and colonic cancer (4) • Motogenic effect via PKC, ERK1/2 and Src in bronchial epithelial cells(12)		(4) Rodrigues et al., 2003a; (12) Graness et al., 2002
TFF3	• Growth inhibition in colon cancer via MAPK/ERK phosphorylation (2) • Resistance to apoptosis via PI3K and EGFR in colon cancer (6) • Resistance to *anoikis* via NF-κB in intestinal epithelial cells (7)	• Cellular invasion independent on EGFR tyrosine kinase activity in kidney and colonic cancer (4) • Cellular invasion by induction of PGE2, PGI2 and TXA-2 in colonic cells (8) • Cell migration via Tyr phosphorylation of β-catenin, activation of EGFR, and modulation of E-cadherin and APC expression in colon cancer (9) • Disruption of cell-substratum contact via FAK phosphorylation in colon cancer (10)	• Angiogenesis by induction of PGE2, PGI2 and TXA-2 in colonic cells (8) • *De novo* angiogenesis via IL-8/CXCR2 and partially via STAT3 in breast cancer cells (11)	(2) Uchino et al., 2000; (4) Rodrigues et al., 2003a; (6) Taupin et al., 2000; (7) Chen et al., 2000; (8) Rodrigues et al., 2001; (9) Liu et al., 1997; (10) Taupin and Podolsky, 2003; (11) Lau et al., 2015

Thus far, all attempts have failed to identify specific TFF receptors and currently no definitive functional membrane-bound receptors exist. Chemokine receptors were proposed to be possible candidates (reviewed in Hoffmann, 2009). However, recombinant dimeric TFF3 has been reported to bind the glycoprotein DMBT1^{GP340}, belonging to the scavenger receptor cysteine-rich (SRCR) superfamily of membrane-bound or secreted proteins, in a Ca^{2+}-dependent manner (Madsen et al., 2013). Interestingly, TFF3 has also been reported to act as a lectin (Reeves et al., 2008), suggesting that it could bind to a plethora of transmembrane glycoproteins.

2. PROJECT'S AIM AND OBJECTIVES

2.1. BACKGROUND AND RATIONALE

The identification of mechanisms involved in the progression from the primary tumor (PT) to metastasis remains a challenge in clinical oncology. At present, research in this field has been based on the hypothesis that a metastatic phenotype is already traceable in the PT. However, favorable and unfavorable prognostic signatures may coexist in different regions of the same tumor, due to spatial heterogeneity, and tumor may adapt and change its molecular features in response to treatment. Thus, to overcome the tumor heterogeneity issue, increasing attention has been paid to circulating tumor cells (CTCs) as they represent an easily accessible source of tumor samples to trace cancer clonal evolution and an early sign of the metastatic process.

CTC studies are particularly useful in breast cancer (BC), known to be heterogeneous, since composed by subsets with distinct features, and able to disseminate at the earliest stages of disease (Hüsemann et al., 2008). Considering the high disseminating potential of BC cells, much attention has been paid to the study of preclinical animal models to shed light on the complexity of BC metastasis biology. Recent studies, in particular, allowed the identification of potential biomarkers, gene signatures and functional mediators for the survival, activation and expansion of metastasis from CTCs in mouse models of many type of tumors (reviewed in Kang and Pantel, 2013), suggesting that mouse models effectively recapitulate the complexity of clinical situations.

Although preclinical research helped deciphering different steps of the metastatic cascade, the clinical validity for many of the identified molecular markers has not been validated yet (reviewed in Bidard et al., 2013), thus suggesting that a refinement of the experimental approaches used for the identification of novel biomarkers is still needed. In the past, the characterization of MDA-MB-231-derived variants showing specific organotropism to lung (Minn et al., 2005), bone (Kang et al., 2003) and brain (Bos et al., 2009), allowed identifying new genetic determinants of the metastatic process able to predict distant metastasis-free survival in BC case series. However, while undoubtedly able to generate

clinically valid information, such BC metastasis experimental models are not suitable for studies on CTC biology as they were obtained by injecting cells directly in the bloodstream.

In the last decade, taking advantage of the improvements in technology for the isolation and molecular analysis of rare cells, efforts were also made to characterize CTCs in animal models. Recent studies provided an experimental demonstration that MDA-MB-231-derived CTCs are more aggressive compared to the parental cell line (Ameri et al., 2010), and that EMT-related changes in cell phenotype are associated with an enhanced ability to intravasate and generate CTCs in MDA-MB-468 xenograft models (Bonnomet et al., 2012). Moreover, CTCs isolated from BC patients were proven to contain a subset of cells showing metastasis-initiating properties when injected in immunocompromised mice (Baccelli et al., 2013).

2.2. HYPOTHESES

Given the unquestionable role of CTCs in tumor progression, we firstly hypothesized that, as tumor cells undergo dynamic changes during hematogenous dissemination, 1) CTCs have a distinct molecular profile compared to solid lesions. Consequently, 2) CTC-specific genes might represent new genetic determinants of the hematogenous phase of metastasis and 3) clinically relevant biomarkers which might help identifying patients at high risk of relapse or progression. Finally, assuming that CTCs are molecularly different from the primary and secondary lesions, it is possible that 4) the expression pattern of CTC-specific genes assessed in the 'liquid biopsy' by real-time monitoring of CTC gene profile has superior prognostic power compared to the molecular profile assessed in the PT.

2.3. AIMS

The aims of this project focused on BC are 1) to provide further knowledge on the molecular mechanisms which regulate cell systemic dissemination, survival in foreign

environments (either when present in ‘liquid phase’ or when colonizing distant organs) and outgrowth at regional/distant sites, by investigating the role of CTC-specific genes in the metastatic cascade, and 2) to identify new prognostic biomarker, by assessing the expression of CTC-specific genes in CTC case series.

2.4. TASKS AND EXPERIMENTAL PLAN

Task #1. Modeling of BC metastases with BC cell lines.

The dissemination and metastatic potential of four BC cell lines (BT-474, MDA-MB-453, MDA-MB-468, MDA-MB-231) was assessed by evaluating CTC load and formation of metastases in different organs, in two independend experiments. The two experiments allowed 1) the identification of an approximate time-point for animal sacrifice, specific for each model, according to the maximum tumor load ethically acceptable and a moderate suffering level per animal, 2) the evaluation of CTC load and the assessment of morphological heterogeneity within the CTC population, and 3) the identification of organs with metastatic involvement, on the basis of information already reported in literature, a macroscopic evaluation of explanted organs at sacrifice, and systematic IHC analysis to assess the frequency and extent of metastases.

Task #2. Molecular characterization of BC xenograft models.

A) CTC-model with the highest CTC load and metastatic potential was characterized by transcriptome analysis of PTs, CTCs, lymph-nodes (LNs) and pulmonary metastases by microarray platform, in two independent experiments.

B) The hypothesis that CTCs have a different molecular profile compared to solid lesions was verified by bioinformatic tools.

Task #3. Identification of candidate genes for basic studies and clinical validation studies.

A) For functional validation studies, genes significantly up-regulated in CTCs compared to the PT and not differentially expressed between CTCs and the parental cells were listed and a panel of candidate genes was selected.

B) For clinical studies, 22 genes of interest were selected. The panel includes i) genes significantly up-regulated in CTCs compared to the solid lesions in the two gene expression profile experiments, ii) genes already proven to be involved in stemness and metastasis in the literature and iii) a classical predictive marker in BC.

Task #4. Assessment of the biological role of candidate genes by *in vitro* and *in vivo* assays.

Loss-of-function experiments were performed by gene knock-down and knock-out approaches and the involvement of one candidate gene in proliferation, migration, invasion, vascular-mimicry, dissemination in blood and metastasis formation at lungs and LNs was assessed by *in vitro* assays and *in vivo* experiments.

Task #5. Comparison of the clinical impact of CTC monitoring by two different methodological approaches.

Blood samples were collected from BC patients with non-metastatic disease, under neoadjuvant therapy, and patients in treatment for metastatic disease. CTCs were analyzed using a positive selection-based (AdnaTest) and a size-based (ScreenCell®) isolation approach before starting therapy and at different time points during treatment. Data were correlated with the clinico-pathological characteristics and follow-up data.

Task #6. Assessment of the clinical relevance of CTC-specific genes in BC CTC case series and gene expression datasets of primary BCs.

A) The expression of the 22 genes selected after the molecular characterization of experimental models was assessed in CTCs isolated by AdnaTest and correlated with the clinico-pathological characteristics and follow-up data.

B) The prognostic relevance of CTC/metastasis-specific genes was also assessed in publicly available datasets.

The project flowchart is illustrated in Figure 2.

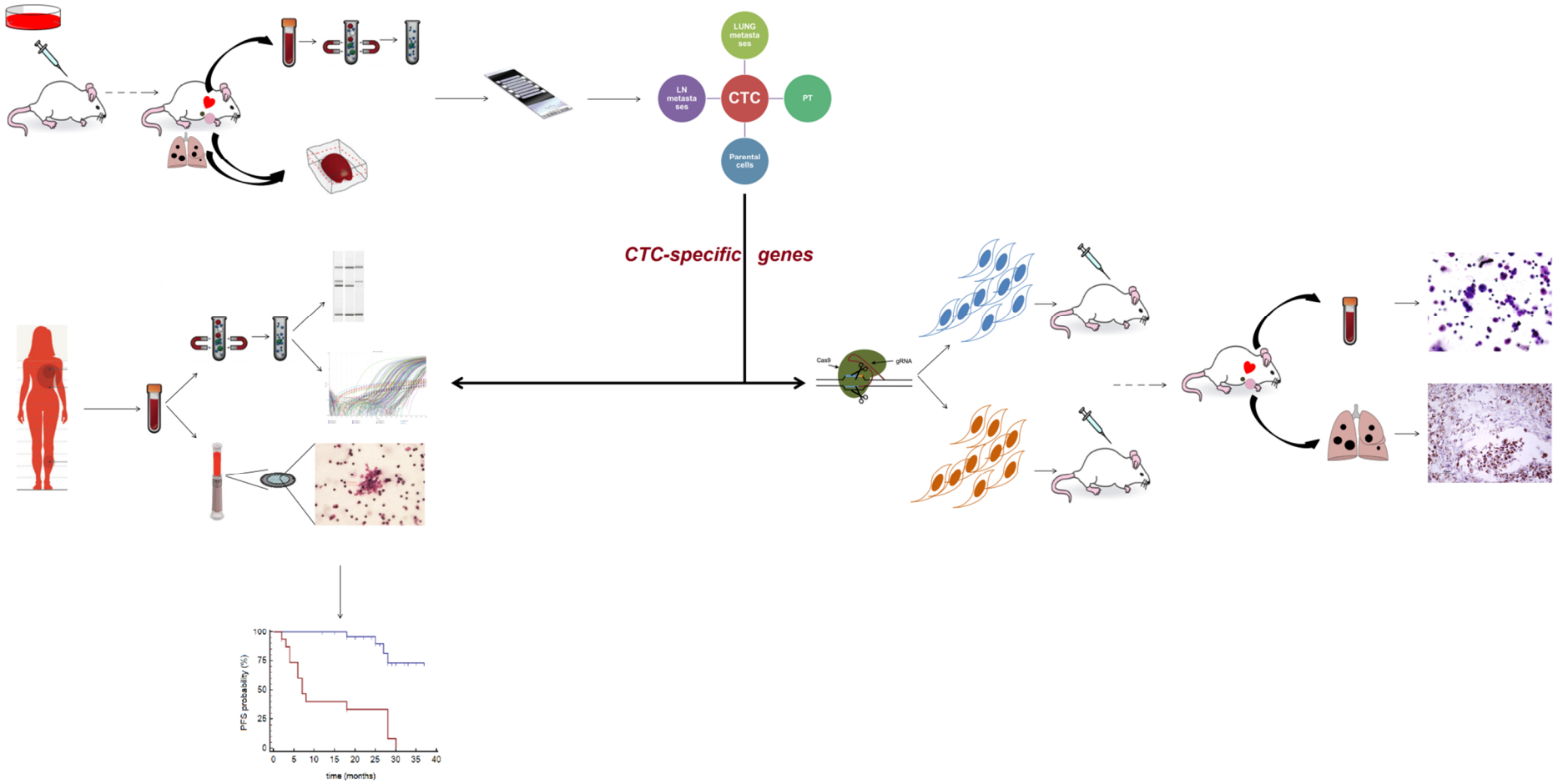


Figure 2. Project flowchart. (***left upper panel***) Breast cancer cells (*Results Section, Chapter 1*) from commercially available cell lines were injected at orthotopic level (*i.e.*, in the mammary fat pad) in immunocompromised female mice. Animals were sacrificed at specific time points according to the experimental model in order to collect fresh blood samples for CTC isolation, fresh bone marrow samples for DTC isolation, primary tumor nodules and organs involved by secondary lesions (*e.g.* lymph-nodes and lung). Microarray analysis (*Results Section, Chapter 2*) was performed to compare the transcriptome of

(***Figure 2. continued***) CTCs with those obtained from DTCs, OCT-embedded frozen sections of primary tumors and metastases, and the parental cell line. Genes significantly up-regulated or down-regulated in CTCs compared to solid lesions were classified as CTC-specific. (***right middle panel***) The biological significance of a candidate CTC-specific gene (*Results Section, Chapter 3*) was assessed by functional validation assays performed *in vitro* (*e.g.* proliferation, migration, invasion, etc.) and *in vivo* (CTC dissemination and metastasis formation, evaluated by cytological or IHC analysis, respectively) using genetically modifies cell lines. (***left lower panel***) In parallel to basic studies, longitudinal monitoring of CTC status and levels (*Results Section, Chapter 4*) in breast cancer case series was carried out using either a positive selection-based method, followed by PCR analysis, or a size-based isolation method, followed by cytological analysis. The clinical significance of a panel of CTC-specific genes (*Results Section, Chapter 5*) was assessed by correlation analysis between each gene expression status, assessed by quantitative PCR, and patients' clinico-pathological features and follow-up data.

3. MATERIALS AND METHODS

3.1. *In vitro* studies

3.1.1. Cell lines

Breast cancer MCF7, BT-474, MDA-MB-453 and MDA-MB-468 cell lines were cultured in Dulbecco's Modified Eagles's Medium (DMEM)/F-12 medium (Lonza, Switzerland) supplemented with 10% South America Fetal Bovine Serum (FBS, Lonza). MDA-MB-231 breast cancer cell line was cultured in DMEM/F-12 medium supplemented with 5% FBS. All cell lines were grown in a controlled chamber at 37°C in a 95% humidified 5% pCO_2 atmosphere. They were purchased from the American Type Culture Collection organization and verified for identity via STR profile analysis using the StemElite™ ID System kit (Promega, Madison, WI, USA). STR profiles yielded a 100% match (all 8 of the 8 loci) in all cases. All experiments were performed using cells from the second to the seventh *in vitro* passage from thawing, showing at least 95% viability by Trypan blue exclusion test.

3.1.2. Cell transfection and transient gene silencing with siRNA molecules

MDA-MB-231 cells were transfected with the *Trans*IT-X2® Dynamic Delivery System (Mirus Bio LLC, Madison, WI, USA) according to the standard protocol provided by the manufacturer. Transfectant reagent toxicity was preliminarily assessed by proliferation assay in a 3-days time-course experiment before performing loss-of-function experiments (growth rate was comparable in transfected and non transfected cells). Delivery efficiency was assessed by transfecting cells with the BLOCK-iT™ Fluorescent Oligo labeled with Cy3 (Thermo Fisher Scientific, Waltham, MA, USA).

Transient *TFF3* silencing was performed using three unique TFF3 (ID 7033) Trilencer-27mer Human siRNA molecules (Origene, Rockville, MD, USA) at concentration of 25 nmol/L. Cells were seeded in 96-well plates to reach 70% confluence (about 2.0×10^4 cells per well). Transfection was performed 24 hours after seeding in a time-course experiment. The Trilencer-27mer Universal Scrambled Negative Control siRNA Duplex

(Origene) was used as control. One out of three tested siRNA molecules was able to reduce TFF3 expression at mRNA (by real-time qPCR) and secreted protein (by ELISA assay) level.

3.1.3. Lentivirus-mediated long term stable gene knock-down

MDA-MB-231 cells were seeded in 6-well plates to reach 70% confluence (about 6.5 x 10^5 cells/well). Infection was performed 24 hours after seeding using the *TFF3* - Human shRNA lentiviral particles, delivering 4 unique 29mer target-specific shRNA and a scramble control by a pGFP-C-shLenti vector (Origene), at a multiplicity of infection (MOI) 5 and 10. Infection efficiency was monitored by cell observation on a Nikon Eclipse TE2000-S fluorescence microscope. After 48 hours medium was changed and cells were put under constant selection with 0.5 µg/mL Puromycin (Sigma-Aldrich, Saint Louis, Missouri, USA) for 10 days. Selected cells were then amplified under standard culture conditions and 72-hours conditioned media were collected to assess protein knock-down (ELISA assay). Two out of the four tested shRNAs were able to produce ~80% TFF3 knock-down compared to control cells.

3.1.4. Gene knock-out by CRISPR-Cas9 system

CRISPR/Cas9-mediated *TFF3* knock-out experiments were performed according to the protocol published by Ran and colleagues (Ran et al., 2013). Guide sequences were designed using the on-line available CRISPR-design tool (http://crispr.mit.edu/). Two sequences designed on Exon 1 and one sequence designed on Exon 2 were chosen (Table 3.1), according to the quality score and mismatches predicted by the algorithm. Guide RNA oligos were cloned in pSpCas9(BB)-2A-Puro (PX459) vectors (Addgene, Cambridge, MA, USA; Figure 3), according to Ran and colleagues' protocol.

MDA-MB-231 cells were seeded in 35 mm culture dishes to reach 80% confluence (about 5.0 x 10^5 cells/well). Cell were transfected 24 hours after seeding with 3 µg of each plasmid or a mixture of two plasmids, at equimolar ratio, to a final amount of 4 µg, using 3 µL of *Trans*IT-X2® Dynamic Delivery System (Mirus) per microgram of DNA. Empty vector was

used as negative control. Cell transfection efficiency was monitored in independent samples, consisting of cells transfected with a pSpCas9(BB)-2A-GFP (PX458) plasmid (Addgene), by cell observation on a Nikon fluorescence microscope. After 48 hours medium was changed and cells were put under constant selection using 0.5 µg/mL Puromycin for 10 days. Selected cells were harvested and cell suspensions were prepared in order to seed 0.6 cells/well. About 2 weeks after seeding, wells were screened to search for proliferating clones, and cell supernatants conditioned for 72 hours were collected from 90 putative *TFF3*^{KO} clones which had reached a 70% confluence.

Table 3.1. Guide sequences for CRISPR/Cas9-mediated *TFF3* knock-out

exon	g#	guide sequence	score	on-target locus
Ex1	#2	**CTGGCAGCCATGACCACCGT (GGG)**	79	chr21: -43735432
Ex1	#5	**CCCCAGCATGCAGAGCGCTC (TGG)**	73	chr21: -43735451
Ex2	#1	**CCCAAGGAGTGCAACAACCG (GGG)**	92	chr21: +43733656

Genomic DNA was extracted by semi-automated approach (QIACube, Qiagen) using the DNeasy Blood & Tissue Kit (Qiagen, Düsseldorf, Germany). DNA concentration was measured by Nanodrop and DNA integrity was assessed by the E-Gel® EX Agarose Gels 1% (Invitrogen™, Thermo Fisher Scientific) using the E-Gel® iBase™ Power System for running and the E-Gel® Safe Imager™ Real-Time Transilluminator for sample migration viewing.

PCR primers were designed to amplify the desired genomic regions in Exon 1 (forward 5'-GACCTCTCCCCTTTGGGAGA-3', reverse 5'-TGCAGAATCCCCCCTTATCC-3') and Exon 2 (forward 5'-CAGCAGTGGTTGAACTCGGC-3', reverse 5'-GGTCCTTGTGCCTCCATCTC-3') using the OligoCalc tool (http://biotools.nubic.northwestern.edu/OligoCalc.html). Optimal thermal profile was set after

testing different annealing temperatures using the Veriti™ Thermal Cycler (Applied Biosystems, Thermo Fisher Scientific).

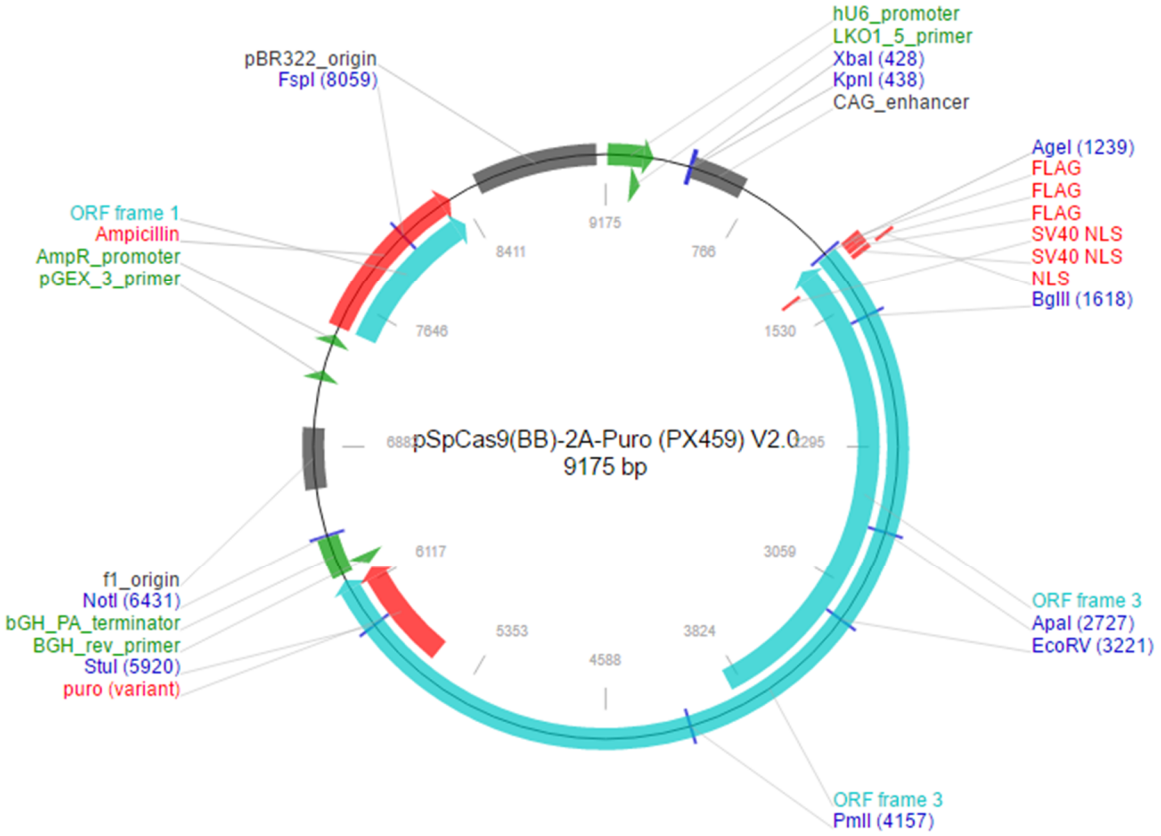


Figure 3. Map of CRISPR/Cas9 vector. Vector backbone PX459, total size 9200 bp. https://www.addgene.org/62988/

The thermal profiles used to amplify the desired genomic regions were: 5 min hold at 95°C, 35 cycles of amplification including a 30 sec denaturation step at 95°C, a 1 min annealing step at 60.4°C or 61.4°C for Exon 1 and Exon 2, respectively, and a 30 sec elongation step at 72°C, and 10 min hold step at 72°C. Fifty nanograms of genomic DNA were amplified using the AmpliTaq Gold® DNA polymerase at a 0.05 U/µL concentration with 2.5 mM $MgCl_2$ (Thermo Fisher Scientific), 0.2 mM concentrated primers and 0.5 mM dNTP mix (Invitrogen™). PCR products were separated and collected using the E-Gel® SizeSelect™ Agarose Gels, 2% (Invitrogen™).

Sanger sequencing was performed at Eurofins (Germany) in two distinct runs in order to sequence both DNA strands using either the forward or the reverse specific primer.

Sequencing output data were analyzed by comparison with a reference sequence (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Possible protein sequences were predicted by the Translate tool (http://web.expasy.org/translate/).

When overlapping sequences were obtained, on the hypothesis that CRISPR/Cas9-mediated genome alteration occurred in heterozygosis, PCR products were cloned in plasmids using the TOPO® TA cloning® Kit for Sequencing (Invitrogen™, Thermo Fisher Scientific) and chemically competent *E. coli* bacteria were transformed. A number of 6 colonies were randomly selected in order to obtain cells expressing one of the two alleles, and for plasmid DNA sequencing was performed. Clones showing gene knock-out in both alleles were included in subsequent studies.

Five out of five clones obtained after transfection with negative control vector showed 100% homology with the reference genomic sequence. Fifteen out of 19 clones called as TFF3-negatìve by ELISA test showed clear gene knock-out at both alleles, by partial gene deletion or single nucleotide insertion/deletion events which caused a frame-shift in the nucleotide sequence, with consequent introduction of a stop codon or change in the predicted aminoacidic sequence. Five validated *TFF3*^{WT} and 8 validated *TFF3*^{KO} clones were used for functional assays.

3.1.5. Proliferation assay

MDA-MB-231 cell proliferation rate was assessed using the CellTiter 96® AQueous One Solution Cell Proliferation Assay (MTS, Promega). For transient silencing experiments, MDA-MB-231 cells were seeded in 24-well plates to reach 70% confluence (about 1.3 x 10^5 cells/well), and transfected for 48 hours. Cells were then seeded in 96-well plates to reach 20% confluence (about 6.5 x 10^3 cells/well) and MTS assay was performed 12, 24, 48 and 72 hours after seeding. Briefly, 20 µL reagent were added to 100 µL culture medium per well and cells were incubated at 37°C in a culture chamber for 2 hours. Cells were lysed by adding 20 µL SDS 10% to stop the reaction, and optical density (O.D.) was immediately measured at a 490 nm absorbance wavelength by the iMark™ Microplate Absorbance

Reader spectrophotometer (Biorad, Hecules, CA, USA). Output O.D. values were obtained by the Microplate Manager 6 software, and corrected for signals emitted by blank (medium without cells). For experiments with MDA-MB-231 cells with *TFF3* stable knock-down, and with *TFF3*^{WT} and *TFF3*^{KO} clones, cells were directly seeded in 96-well plate and MTS assay was performed as described before.

3.1.6. Migration assay

MDA-MB-231 migration ability was assessed using a Boyden chamber assay. For transient *TFF3* silencing experiments, cells were seeded in 24-well plates to reach 70% confluence, and transfected with the *TFF3* siRNA molecule for 24 hours. Medium was removed and replaced by serum-free medium for cell starvation. After 24 hours 6.0×10^4 cells were seeded in serum-free medium on 6.5-mm diameter Transwell® membrane cell culture inserts with 8.0-µm pore diameter (Corning®, New York, USA) and migration was stimulated by using complete growth medium as an attractant in the bottom well. Cells on the top of the membrane were removed 24-36 hours after seeding, using a cotton-tipped applicator. Cells were washed in Dulbecco's Phosphate-Buffered Saline (DPBS, Lonza), fixed in ethanol absolute for 20 minutes at -20°C and stained with 0.4% sulforhodamine B (SRB, Sigma-Aldrich) for 30 minutes at room temperature (RT). Migration areas were acquired on a Nikon microscope and densitometric analysis was performed using the ImageJ software version 1.51g (https://imagej.nih.gov/ij/).

For experiments with MDA-MB-231 cells with *TFF3* stable knock-down, and with *TFF3*^{WT} and *TFF3*^{KO} clones, cells underwent starvation for 24 hours and seeded for migration assay as described before. Migrating GFP-expressing cells (stable *TFF3* knock-down) were directly visualized and images were acquired on a Nikon fluorescence microscope, whereas *TFF3*^{WT} and *TFF3*^{KO} clones were stained with SRB as described before.

3.1.7. Invasion assay

MDA-MB-231 invasion ability was assessed using a Boyden chamber assay. MDA-MB-231 cells with *TFF3* stable knock-down, and *TFF3*^{WT} and *TFF3*^{KO} clones were seeded in 24-well plates to reach 70% confluence for 24 hours. Medium was removed and replaced by serum-free medium for cell starvation. After 24 hours 6.0×10^4 cells were seeded in serum-free medium on 6.5 mm diameter Transwell® membrane cell culture inserts with 8.0-µm pore diameter membranes (Corning) coated with BD Matrigel™ Basement Membrane Matrix Growth Factor Reduced (BD Biosciences, San Jose, CA, USA) at a 200 µg/mL concentration, and invasion was stimulated by using complete growth medium as an attractant in the bottom well. Matrigel and cells on the top of the membrane were removed 24-36 hours after seeding, using a cotton-tipped applicator.

Invading GFP-expressing cells (stable *TFF3* knock-down) were visualized and images were acquired on a Nikon fluorescence microscope; densitometric analysis was performed using the ImageJ software.

For experiments with *TFF3*^{WT} and *TFF3*^{KO} clones, invading cells were stained using the SRB protocol.

3.1.8. Vascular-mimicry assay

Vascular-mimicry ability was assessed using MDA-MB-231 cells with stable *TFF3* knock-down. Cells suspensions at 4.0×10^5 cells/ml density were prepared in serum-free medium and 2.0×10^4 cells were seeded onto pure Corning® Matrigel® Basement Membrane Matrix (Corning) coated 96-well plates (35 µL per well). Cells were monitored for 6-12 hours to assess loop formation, and images were acquired on a Nikon TE 2000-S fluorescence inverted microscope.

3.1.9. Cell spiking-capture experiments

Spiking experiments were performed by manual pipetting of defined numbers of cells, showing at least 99% viability by Trypan blue exclusion assay, into 5 mL whole blood

samples withdrawn from healthy donors. Blood samples were processed within 1 hour from spiking, using the AdnaTest EMT-1/Stem CellSelect or EMT-2/Stem CellSelect kits (AdnaGen, Langenhagen, Germany). Tumor cell detection was performed using the AdnaTest EMT-1/Stem CellDetect kit (AdnaGen), according to the manufacturer's instructions. Primer mix (forward and reverse) and experimental protocol for *MET* amplification were kindly provided by AdnaGen. Briefly, 4.0 µL of cDNA were amplified using the HotStarTaqMaster Mix (Qiagen, Düsseldorf, Germany) in a final volume of 25.0 µL, using the following PCR thermal profile: hold step at 95°C for 15 min, 35 cycles of denaturation, annealing and amplification at 94°C for 60 seconds, 58°C for 2 minutes and 72°C for 30 seconds, respectively, and hold step at 72°C for 5 minutes.

3.2. Animal studies

3.2.1. Mouse models

Animal experiments were performed according to Italian law (D.L. 116/92 and following additions), which enforces the 2010/63/EU Directive; the study was approved by the Ethical Committee for Animal Experimentation at Fondazione IRCCS Istituto Nazionale dei Tumori, Milan (INT). All efforts were deployed to minimize animal suffering, following the most recently published version of recommended guidelines (Workman et al., 2010).

Six- to 8-week-old female NOD.CB17-*Prkdc*^{scid}/J (NOD scid) and NOD.*Cg-Prkdc*^{scid} *Il2rg*^{tm1Wjl}/SzJ (NSG) mice were purchased from Charles River (Wilmington, MA, USA) and The Jackson Laboratory (Sacramento, CA, USA), respectively. Animals were bred in individually ventilated cages, 3 to 5 animals per cage.

Mice were anesthetized by intraperitoneal injection of ketamine (100 mg/kg) and xylazine (5 mg/kg) cocktail before tumor cell orthotopic injection and sacrifice. Animals were sacrificed by cervical dislocation at the desired experimental time point or immediately when

they started to show signs of moderate level of suffering (e.g. decrease in activity, hunched appearance, ruffled hair coat, respiratory distress).

3.2.2. Tumor cell injection

Cell lines were tested for Mycoplasma contamination by the MycoAlert® mycoplasma detection kit on the GloMax® 20/20 luminometer before performing injection in mice. Tumor implant was performed in 9-10 week-old anesthetized mice by injection with a 30G needle syringe in the mammary fat pad (m.f.p.). Intravenous injection was performed using 10^6 or 2 x 10^6 cells, resuspended in 400 µL DPBS.

The experimental scheme of each xenograft model is reported in Table 3.2. ECM Gel Matrigel was purchased from Sigma-Aldrich. Mice injected with BT-474 cells were treated with estrogens by subcutaneous implantation of 0.72 mg 90-day release 17-β-estradiol pellets, performed using a trocar, 24 hours before tumor implant.

Table 3.2. Experimental conditions for CTC modeling with BC cell lines

Cell line	*Mouse model*	*Number of cells per injection*	*Cell suspension composition and volume*	*Injection sites*	*Hormone supplementation*
BT-474	NSG	5.0 x 10^6	1:1 DPBS:Matrigel, 80 µL	4^{th} left mammary gland	yes
MDA-MB-453	NSG	10^7	1:1 DPBS:Matrigel, 90 µL	4^{th} left mammary gland	no
MDA-MB-468	NSG	5.0 x 10^6	1:1 DPBS:Matrigel, 80 µL	2^{nd} right and 4^{th} left mammary glands	no
MDA-MB-231	NOD scid	5.0 x 10^6	1:1 DPBS:Matrigel, 80 µL	2^{nd} right and 4^{th} left mammary glands	no

For the time-course experiments with MDA-MB-231 cells, 4 groups of 6 animals each were injected with cells according to the standard scheme (Table 3.2). Animals were

randomized before sacrifice at the defined time points (day 35, 50, 65 and 80) according to the tumor growth rate and the cage where they had been bred.

Pools of MDA-MB-231 *TFF3*^{KO} or *TFF3*^{WT} clones were prepared after mixing equal amounts of cells to reach 5.0 x 10^6 cell suspensions. Injection was performed according to the standard scheme for MDA-MB-231 reported in Table 3.2.

Tumor growth was monitored every week using a caliper and tumor weight (TW) was estimated by the TW (g) = (D x d^2)/2 formula, where D and d represent the longest and the shortest diameter, respectively, of the tumor.

3.2.3. Collection of tissues and organs

Blood samples were collected in EDTA (1.8 mg/mL) tubes by cardiac puncture in anesthetized mice by using a 1 mL 26G needle syringe. Mice were immediately sacrificed and primary tumor nodules and organs (lung, axillary or inguinal lymph-nodes, ovaries, liver, kidneys, brain) were collected and fixed in 10% neutral buffered formalin (Bio-Optica, Milan, Italy) for 18 to 24 hours. Samples were then embedded in paraffin (FFPE).

For gene expression profile experiments with MDA-MB-231 xenografts, tumors, lungs, and axillary or inguinal lymph-nodes were OCT-embedded and snap-frozen in liquid nitrogen. Samples were stored at -80°C until processing.

3.2.4. Circulating tumor cell isolation

For gene expression analyses, CTCs were isolated using the AdnaTest EMT-2/Stem CellSelect kit and the ScreenCell® Molecular biology kit. For AdnaTest, blood samples were incubated for 25 minutes with 50 µL of immunomagnetic beads and captured cells were isolated using a magnet (AdnaMag magnetic separator), according to the manufacturer's instructions. Residual blood was filtered by the ScreenCell® Molecular Biology kit, according to the manufacturer's instructions. Cell lysates were obtained by both samples using the AdnaGen proprietary lysis buffer and samples were stored at -20°C for no more than 14 days until processing.

For *in situ* analyses, CTCs were isolated using the ScreenCell® Cyto kit (ScreenCell, Sarcelles, France), according to the manufacturer instructions. The isolation supports (IS) were stained with Hematoxylin Solution S (Merck, Darmstadt, Germany) for 1 minute and Shandon Eosin Y Aqueous Solution (Thermo Fisher Scientific Inc., Waltham, MA, USA) for 30 seconds, or with pure May-Grünwald solution for 2.5 minutes, followed by a 2.5-minute incubation step with May-Grünwald solution diluted 1:2 with distilled water pH 7, and Giemsa solution (Merck) 1:10 diluted in distilled water pH 7, for 10 minutes.

Cells disseminated in the bone marrow were obtained by flushing of femurs with DPBS and they were isolated using the AdnaTest and ScreenCell approaches for nucleic acids analysis, as described above.

3.2.5. Tumor cell detection and quantification

For CTC quantification, mRNA was isolated from cell lysates using the AdnaTest EMT-2/Stem CellDetect kit and resuspended in 10 µL nuclease-free water. Half RNA volume was stored at -80°C for microarray gene expression profile analyses, while the remaining was retrotranscribed using the High-Capacity cDNA Reverse Transcription Kit with RNase Inhibitor (Applied Biosystems™, Thermo Fisher Scientific) in a 10 µL reaction volume. Two out of 10 µL cDNA were amplified for indirect quantification, which was performed using a standard curve constructed by correlating different numbers of MDA-MB-231 cells (2-5-10-20-40-100-200-1,000-5,000-10,000-50,000-100,000), directly spiked in the lysis buffer by manual pipetting or after serial dilution from cell suspensions (when ≥200 cells), to C_t values obtained by qPCR using a TaqMan assay specific for human *ACTB*.

Total RNA from MDA-MB-231 cell pellets or from two 10-µm thick frozen sections of tumors, lungs and lymph-nodes OCT-embedded samples was extracted by the Agencourt® RNAdvance™ Tissue Kit (Beckman Coulter, Breas, CA, USA), according to the manufacturer's instructions, and quantified at Nanodrop. RNA quality was assessed by the Agilent RNA 6000 Nano Kit (Agilent Technologies, Santa Clara, CA, USA). RIN values ranged from 9.5 to 9.9 for cultured cells and from 6.7 to 9.3 for frozen sections. Convenient

RNA volumes were retrotranscribed using the High-Capacity cDNA Reverse Transcription Kit with RNase Inhibitor in 20 µL reaction volume, and tumor cells were quantified by qPCR as described above.

3.3. Clinical studies

3.3.1. Study design and case series

Patients with histologically confirmed breast cancer and no evidence of metastatic disease (M0) who were receiving anthracycline/taxane-based neoadjuvant therapy and trastuzumab if HER2-positive, and breast cancer patients with metastatic disease (M+) who were going to start a new line of systemic treatment (mostly endocrine treatment) were recruited at the Department of Medical Oncology at INT. Fifty-nine M0 and 56 M+ patients were enrolled from February 2011 to September 2016.

For each patient, samples were collected at baseline (before starting treatment) and during treatment at established times: 3 and 6 months after initiation of neoadjuvant treatment and around 4 weeks after mastectomy in M0 cases; at the beginning of a new line of treatment and at 3 months from treatment start or at progression in M+ cases.

Thirty female healthy subjects donated blood samples for technical fesibility experiments and to define CTC positivity cut-off values.

This study was submitted and approved by the Institutional Review Board and the Ethics Committee at INT. All patients and healthy subjects signed and informed consent for withdrawing blood samples dedicated to this research.

3.3.2. Blood sampling

Samples of peripheral venous whole blood were drawn from BC patients and healthy donors using a 26 G needle and collected in K_3EDTA or K_2EDTA BD Vacutainer tubes, if

processed with the AdnaTest (AdnaGen, AG, Langenhagen, Germany) or ScreenCell® Cyto (ScreenCell, Sarcelles, France) kits, respectively. The first blood tube was discarded to minimize the risk of contamination with skin epithelial cells. Samples were stored in the dark at 4°C and processed within 1 hour from withdrawal according to the manufacturer's instructions.

3.3.3. CTC enrichment and detection by AdnaTest kits

CTC enrichment with the AdnaTest EMT-1/Stem CellSelect and EMT-2/Stem CellSelect kits was performed in parallel on two 5-mL blood samples from the same patient according to the manufacturer's instructions.

For CTC detection, the expression levels of *EPCAM*, *MUC1*, *ERBB2*, *PIK3CA*, *AKT2*, *TWIST1* and *ALDH1A* transcripts were evaluated by semi-quantitative multiplex PCR according to the manufacturer's instructions and using the suggested thresholds for positivity (0.15 ng/µL for *EPCAM*, *MUC1*, *ERBB2* and *ALDH1*, and 0.25 ng/µL for *PIK3CA*, *AKT2* and *TWIST1*). According to the results obtained in a series of 5 healthy donors, samples with *ACTB* expression <0.70 ng/µL were excluded from the analysis, whereas those passing the quality control where considered as positive when positive for at least 1 of the above mentioned markers.

3.3.4. CTC isolation and detection by ScreenCell® Cyto kits

All blood samples were processed within 1 hour after collection using the ScreenCell® Cyto kit (ScreenCell, Paris, France) according to the manufacturer's instructions. Briefly, blood was diluted in 4 mL of red blood cell lysis and fixation buffer and incubated for 8 minutes at room temperature. Three filtrations of 3-mL blood samples each were separately performed for each patient; IS were rinsed with DPBS, collected from the device, air-dried and immediately stained at room temperature for 1 minute with Hematoxylin Solution S (Merck, Darmstadt, Germany) and then for 30 seconds with Shandon Eosin Y

Aqueous Solution (Thermo Fisher Scientific Inc., Waltham, MA, USA). IS were stored at -20°C until cytological evaluation.

All samples were analyzed by the same pathologist without knowledge of the clinical data. Major criteria for CTC identification were a high nucleus-to-cytoplasm ratio (≥0.75) and large nuclear size (≥20 µm), whereas minor criteria included irregular nuclear contours and nuclear hyperchromatism. The cytomorphological analysis and CTC count were based on the previously reported criteria of malignancy (Hofman et al., 2011). Circulating tumor microemboli (CTMs) were defined as clusters of at least 2 CTCs, often mixed with platelets and various leukocytes, showing criteria of malignancy like those described for single CTCs. The nucleus-to-cytoplasm ratios in single CTCs are similar to those in CTC aggregates (Cho et al., 2012). Results were expressed as numbers of CTCs and CTMs for single membranes.

For each patient, total CTC or CTM numbers derived from 3 membranes (corresponding to 9 mL of blood) were added together to better meet the criteria for accurate detection of rare events following the Poisson probability distribution (Allan and Keeney, 2010). Membranes showing poor quality of cytology, estimated on the basis of poor preservation of the leukocytes, were excluded from the analysis. Samples were rated as CTC or CTM positive if at least 1 CTC or CTM was detected in the 3 membranes.

3.4. Gene expression studies

3.4.1. RNA extraction, quantification and quality control testing

Total RNA was isolated from cell pellets and OCT-embedded frozen sections using the Agencourt® RNAdvance™ Cell v2 or the Agencourt® RNAdvance™ Tissue Kit (Beckman Coulter) according to the manufacturer's instructions. The procedure is based on Solid Phase Reversible Immobilization paramagnetic bead technology, which ensures efficient removal of genomic DNA and other contaminants. RNA quantification and integrity control

were performed by Nanodrop and the Agilent RNA 6000 Nano Kit (Agilent Technologies) by the Bioanalyzer, respectively. RIN values were higher than 9.5 for cell lines and ranged from roughly 6.5 to 9 for frozen animal tissues.

Messenger RNA was isolated using the AdnaTest EMT-1/Stem CellDetect kit, containing (dT)25-coated Dynabeads® for the isolation of pure polyadenylated RNA.

Isolated RNA was eluted in convenient volumes of nuclease-free water (Ambion, Thermo Fisher Scientific) and stored at -80°C until processing.

3.4.2. cDNA synthesis and quantitative real-time PCR

cDNA synthesis was performed using the High-Capacity cDNA Reverse Transcription Kit with RNase Inhibitor (Applied Biosystems™, Thermo Fisher Scientific) in a GeneAmp® PCR System 9700 thermal cycler, using a modified thermal profile (10 minutes at 25°C, 60 minutes at 42°C, 5 minutes at 85°C). Reverse transcription products were stored at -20°C until processing.

Amplification of specific genes was performed by real-time qPCR using the single TaqMan assays listed below:

target	*assay ID*
Human *ACTB*	Hs03023943_g1
Human *ELF3*	Hs00963882_g1
Human *GAPDH*	Hs00266705_g1
Human *NR4A1*	Hs00374230_m1
Human *TFF1*	Hs00907239_m1
Human *TFF2*	Hs00193719_m1
Human *TFF3*	Hs00902278_m1
Murine *GAPDH*	Mm99999915_g1

Ten nanograms of cDNA were amplified in technical triplicates using the TaqMan® Universal PCR Master Mix, according to the standard protocol. PCR reactions were run in MicroAmp Fast Optical 96-Well Reaction Plates by the QuantStudio™ 12K Flex Real-Time PCR System, according to the standard thermal profile. The threshold cycle (C_t) for each

gene was automatically set by the 7900 SDS v2.4 dedicated software. Analysis of relative gene expression was performed applying the $2^{-\Delta\Delta CT}$ method. *GAPDH* was used as endogenous control gene for qPCR analysis to assess *TFF3* silencing upon siRNA delivery.

3.4.3. Low-density array gene expression profiling

Gene expression profiling of CTCs from clinical samples was performed using custom 384-Well Microfluidic Card TaqMan® Gene Expression Assays for the analysis of 23 genes and an endogenous control in technical duplicate. Target specific pre-amplification was performed using a Custom TaqMan® PreAmp Pool of 24 TaqMan assays, according to the manufacturer's instructions.

Table 3.3. Low-density array gene expression assays and relative positivity cut-off C_q values

target gene	*assay ID*	*cut-off*
NR4A1	Hs00374226_m1	29.90
HDAC10	Hs00368899_m1	27.17
CRIP1	Hs00832816_g1	24.09
TPPP	Hs00389316_m1	40.00
GAS2L1	Hs00977983_g1	29.38
FIS1	Hs00211420_m1	27.21
FCF1	Hs01056861_gH	40.00
STRN4	Hs01026676_m1	29.13
TAF6	Hs00425763_m1	29.18
TFF1	Hs00907239_m1	40.00
TFF2	Hs00193719_m1	40.00
TFF3	Hs00902278_m1	40.00
ELF3	Hs00963881_m1	40.00
ESR1	Hs00174860_m1	31.00
ADPRHL1	Hs00293405_m1	40.00
KLC2	Hs03988192_m1	28.53
PTPRC	Hs04189704_m1	-
GIGYF1	Hs01119153_g1	29.01
YAP1	Hs00902712_g1	40.00
OPN (SPP1)	Hs00959010_m1	40.00
PTPN11	Hs01590331_gH	29.50
SPARC	Hs00234160_m1	31.51
SOX2	Hs01053049_s1	40.00
GAPDH	Hs02758991_g1	-

PCR were run in technical duplicate by the 7900HT Fast Real-Time PCR System. Data were analyzed using the Relative Quantification analysis module RQ Manager version 2.3. The relative threshold method (C_{rt}) was used to set the threshold cycle, as it was proven to be more robust than the C_t baseline threshold method when analyzing multiple gene expression data from low-density arrays. Indeed, the C_{rt} method accounts for low reaction volumes and associated differences in fluorescence levels. Two quality control parameters were set before performing C_{rt} analysis: the AMP score (rejection threshold <1.0) and the Calculated confidence in the Cq value (rejection threshold <0.8), both related to the performance of the amplification, scored according to an ideal sigmoidal amplification curve.

Mean equivalent C_q values obtained after C_{rt} were used for subsequent gene expression analyses. Positivity threshold for each target gene was set according to the mean equivalent C_q obtained from 5-mL blood samples of 6 female healthy donors, processed by the AdnaTest EMT-2/Stem CellSelect kit. A gene was considered as expressed in a CTC sample if the C_q value was lower than the lowest C_q observed among control individuals. Overall mean±SD C_q values for *GAPDH* and *PTPRC* in patients' samples were 33.94±5.63 and 34.22±3.64, respectively. Samples showing $C_{q(GAPDH)}$ = 40 and $C_{q(PTPRC)}$ = 40 were excluded from the analysis. Target genes and relative TaqMan assays and positivity cut-off C_q values are listed in Table 3.3.

For bioinformatic analyses, data were normalized on the geometric mean of C_q values of *GAPDH* and *PTPRC* control genes.

3.4.4. Microarray gene expression profiling

RNA samples from the MDA-MB-231 xenograft model, each corresponding to 5,000 tumor cells according to the indirect quantification approach, and the parental cells (50 ng total RNA) were profiled using the Illumina Human Whole-Genome DASL HT Assay (Illumina, Inc., San Diego, CA, USA), according to the manufacturer's instructions RNA is converted to cDNA using biotinylated oligo(dT) and random nonamer primers. The biotinylated cDNA is then annealed to the DASL Assay Pool (DAP) probe groups. Probe

groups contain oligonucleotides specifically designed to interrogate each target sequence in the transcripts. These probes span about 50 bases, making it possible to profile partially degraded RNA.

Universal human reference RNA from Stratagene (San Diego, CA, USA) and Mouse Universal Reference from Clontech (Mountain View, CA, USA) were used as positive and species-specificity controls, respectively (50 ng each). The BeadChips were imaged on the BeadArray Reader. Illumina BeadScan software was used for image acquisition.

3.5. Protein analyses

3.5.1. Measurement of secreted proteins

Conditioned medium was centrifuged at 1,000 RCF for 5 minutes at 4°C to remove floating cells and debris, and stored at -80°C in sub-aliquots. Secreted *TFF3* concentration was measured by the Human TFF3 Quantikine® ELISA Kit (R&D systems, Minneapolis, MN, USA), according to the manufacturer's instructions. The assay was verified by the manufacturer for not interfering with TFF1, TFF2 and other available related molecules. TFF3 concentrations (pg/mL) were calculated using a standard curve.

3.5.2. Immunohistochemical staining

Four-micron thick FFPE sections from tumor nodules, lungs, lymph-nodes, ovaries, liver, brain and kidneys collected from experimental models were stained using a monoclonal antibody against human COX IV (clone 3E11, isotype IgG, produced in rabbit) purchased from Cell Signaling Technology (Danvers, MA, USA). Antigen retrieval was performed at 95°C for 30 minutes in Sodium Citrate Buffer (10 mM Sodium Citrate, 0.05% Tween 20, pH 6.0). Endogenous biotin blocking was performed for liver sections only, using the Dako Cytomation Biotin Blocking System (Dako, Troy, MI, USA). Samples were incubated with the primary antibody at a 1:1,000 dilution at 4°C overnight.

Four-micron thick frozen sections from tumor nodules, lung and lymph-nodes were fixed in 10% neutral buffered formalin (Bio-Optica) for 10 minutes at RT and incubated with the antibody for 1 hour at RT.

Antibody visualization was obtained using the EnVision®+ System- HRP Labelled Polymer (Dako). Nuclei were counterstained with Mayer hematoxylin (Bio-Optica).

3.6. Bioinformatic analyses

Raw data from microarray analysis were pre-processed using the R/Bioconductor package “lumi” that provides statistical methods for analysis of Illumina microarray data. As first step of pre-processing data were log(base2)-transformed, as the distribution of intensity values in the linear scale is skewed towards the background intensity. Log transformation makes the data distributions closer to a Normal distribution and makes the variance independent of the mean. After log-transformation data were normalized using the robust spline normalization method (Du et al., 2008). Probes not associated to HUGO gene symbols were filtered out. Multiple probes mapping to the same gene were collapsed selecting the probe with the highest inter-quartile range across samples.

Differential expression analysis was performed using the “limma” package (Smyth, 2004) that combines linear models with moderated t-statistic to identify differentially expressed genes across experimental conditions. The moderated t-statistics has the same interpretation as an ordinary t-statistic except that the standard errors are moderated across genes, *i.e.*, queezed towards a common value, using a simple Bayesian model. P-values obtained from limma were adjusted for multiple-testing using the Benjamini-Hochberg false discovery rate (FDR) to reduce the number of false positives. Genes showing an absolute fold change ≥ 2 and an FDR <0.05 were considered significantly differentially expressed.

Principal variance component analysis (PVCA), used to assign the percentage of variability to known biological and technical factors present in gene expression data, was performed using the “pvca” package. This approach combines the strengths of principal

component analysis to reduce data dimension and variance component analysis to fit a mixed linear model using factors of interest as random effects to estimate and partition the total variability. This method allows determining which sources of biological and technical variability are more prominent in a microarray dataset.

Unsupervised hierarchical clustering was performed using average linkage and 1 - Pearson's correlation as distance measure. Clustering techniques allow to group samples in a given dataset according to the similarity in their gene expression profiles.

Gene Ontology (GO) analysis was carried out using the "topGO" package on the lists of genes found significantly differentially expressed in both experiments in order to assess the over-representation of genes belonging to specific biological processes. To summarize the results of GO analysis the overlap coefficient between each pair of significant GO terms was calculated as a measure of similarity, and terms with an overlap coefficient ≥0.5 were visualized as a network using Cytoscape version 3.2.1 and manually annotated.

Gene expression datasets of lymph node-negative primary breast tumors were retrieved from NCBI GEO repository with accession numbers GSE2034, GSE2990, GSE5327, GSE7390 and GSE11121. All downloaded datasets were generated on the Affymetrix GeneChip® Human Genome U133A array and raw CEL files were pre-processed using the frozen RMA (fRMA) algorithm that enables the analysis of Affymetrix data at the single sample level (McCall et al., 2010).

3.7. Statistical analyses

Data were analyzed using conventional descriptive statistics. Contingency tables were analyzed by two-tailed Fisher's exact test, two-tailed χ^2 test with Yates'correction and by χ^2 test, when appropriate, and the quantitative measure of the agreement between categorical variables was evaluated by Cohen's kappa statistics, κ. Differences among ordinal variables were assessed by Mann-Whitney's *U* or by Student's *t* test or by ANOVA (moderated t-statistics).

The Kaplan-Meier method was used for estimation of time-to-event outcomes such as progression-free survival (PFS), as a function of overall CTC positivity or CTC-specific gene status (negative or positive), or distant metastasis-free survival (DMFS), as a function of CTC-specific gene expression level (low, intermediate or high), in CTC case series collected at INT and in publicly available BC datasets. Cox proportional hazards regression was used to investigate the prognostic role of CTC molecular profile, and of the other clinico-pathological or CTC status-related factors on PFS, with relative hazard ratios (HR) and 95% confidence intervals (CI), whereas log-rank test was used to investigate the prognostic role of CTC-specific genes in publicly available BC datasets and of the CTC status assessed by AdnaTest in CTC case series collected at INT.

All tests for comparison of experimental groups in basic studies were performed using the GraphPad Prism or R software. All tests for clinical correlation analyses were performed using the SAS software, version 9.2. All tests were two-sided and $P \leq 0.05$ was adopted as the significance threshold.

4. RESULTS

4.1. Modeling dissemination and metastasis formation with breast cancer cell lines

4.1.1. Detection of CTCs and metastases in xenograft mouse models

Direct visualization of disseminated tumor cells is fundamental to measure the metastatic burden and to analyze morphological parameters related to tumor cell heterogeneity, both in blood and solid tissues. For these reasons, CTCs and metastases in BC experimental models obtained by orthotopic xenograft of commercially available cell lines were detected using cytological and immunohistochemical approaches.

To this aim, CTC enrichment, detection and quantification were performed using a size-based isolation method, which physically excludes the majority of hematopoietic cells by exploiting the larger diameter of tumor cells, coupled to a cytological analysis by a referral pathologist. This approach also facilitates the identification of different subpopulations of CTCs, either presenting as single cells or clusters or circulating tumor microemboli (CTM), whose representative images were reported in Figure 4.1.1. Generally, CTCs are characterized by i) a larger nucleus (generally 13 to 15 μm in diameter) compared to leukocytes, which appear hardly larger (about 7-8 μm in diameter) than membrane pores (6.5±0.33 μm), ii) a high nucleus-to-cytoplasm ratio (>0.5), iii) a dense basophilic and irregularly outlined nucleus and iv) a pale-bluish ring of cytoplasm, forming a thin rim encircling the nucleus. CTC clusters are defined as groups of two or more CTCs, sometimes mixed with platelets (*i.e.*, CTM).

Given the weak metastatic ability of some BC cell lines, direct inspection of biological samples to find single disseminated cells or small foci was performed by IHC analysis using a commercially available antibody against human COX IV, suggested by the manufacturer as non-crossreactive against *Mus musculus* species' antigens. The species-specificity of anti-COX IV antibody was assessed on a series of 4 FFPE sections from different organs of 3 non-tumor-bearing NOD scid mice. Despite low-intensity scattered signals were observed

at nervous tissue level, sections from lung, liver, lymph-node and kidney did not display reactivity after IHC procedure (Figure 4.1.2), thus indicating that the antibody is highly specific and that the approach is adequate to assess the presence and extent of metastatic foci without turning to the pathologist for consultation.

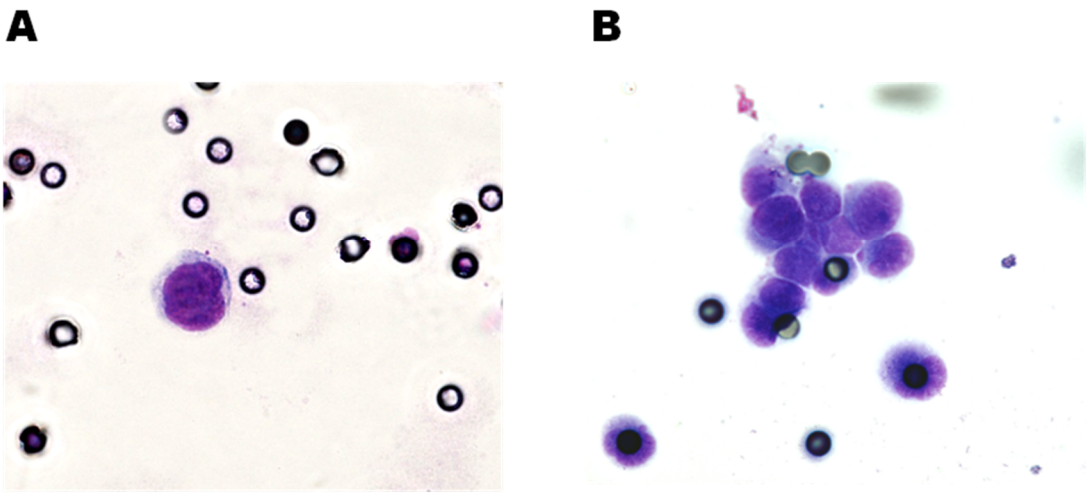


Figure 4.1.1 Single CTCs and CTC cluster from experimental breast cancer models. Images (600X total magnification, oil immersion) of cytological samples (May-Grünwald-Giemsa staining) obtained by ScreenCell® Cyto kits represent (**A**) 1 CTC from MDA-MB-453, and (**B**) 1 cluster of CTCs mixed with some platelets, and 2 single and separated CTCs (below the cluster), from MDA-MB-231 xenograft mouse models.

4.1.2. Hematogenous dissemination and metastatic potential of BC cell lines

BT-474, MDA-MB-453, MDA-MB-468 and MDA-MB-231 cells, belonging to the HER2-positive ER-positive, HER2-positive ER-negative, basal A and basal B molecular subtypes respectively, were injected in the m.f.p. of immunocompromised mice and monitored for tumor growth (Figure 4.1.3), CTC release (Figure 4.1.4 and Table 4.1.1) and invasion of lymph-nodes and distant organs (Table 4.1.2) in two independent experiments (hereafter referred to as “Exp1” and “Exp2”).

Consistently with the fast tumor growth rate, MDA-MB-231 xenograft animals were sacrificed around day 80 after injection, whereas animals injected with BT-474, MDA-MB-453 and MDA-MB-468 cell lines, whose tumor growth rate was lower than MDA-MB-231 cells, were sacrificed at later time points (after 100 days) (Figure 4.1.3).

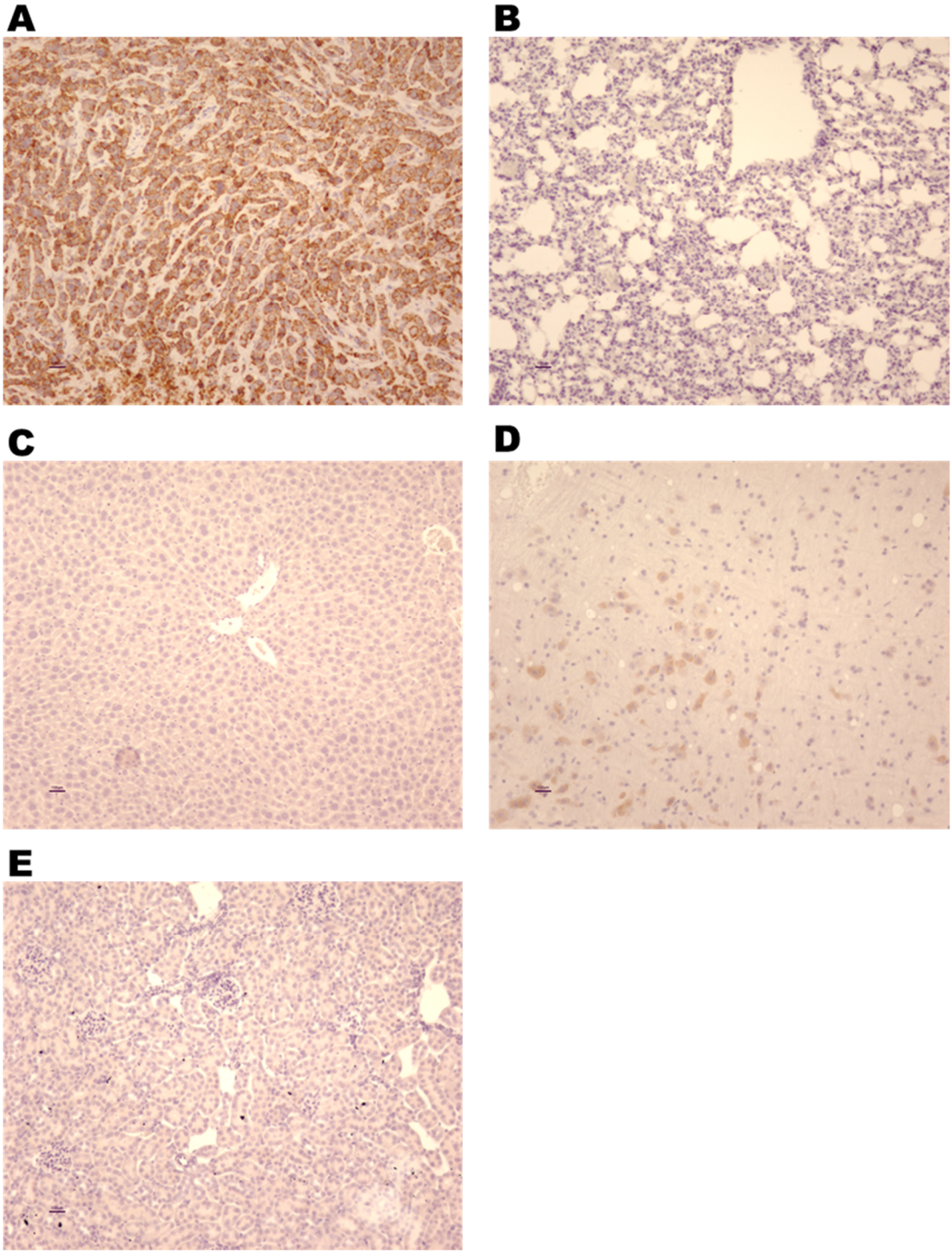


Figure 4.1.2. Detection of tumor cells by human COX IV IHC analysis. Images (100X total magnification) of COX IV stained FFPE sections from (**A**) a primary tumor nodule obtained from MDA-MB-231 xenograft model, and from (**B**) lungs, (**C**) liver, (**D**) brain and (**E**) kidneys collected from a non-tumor-bearing immunocompromised mouse.

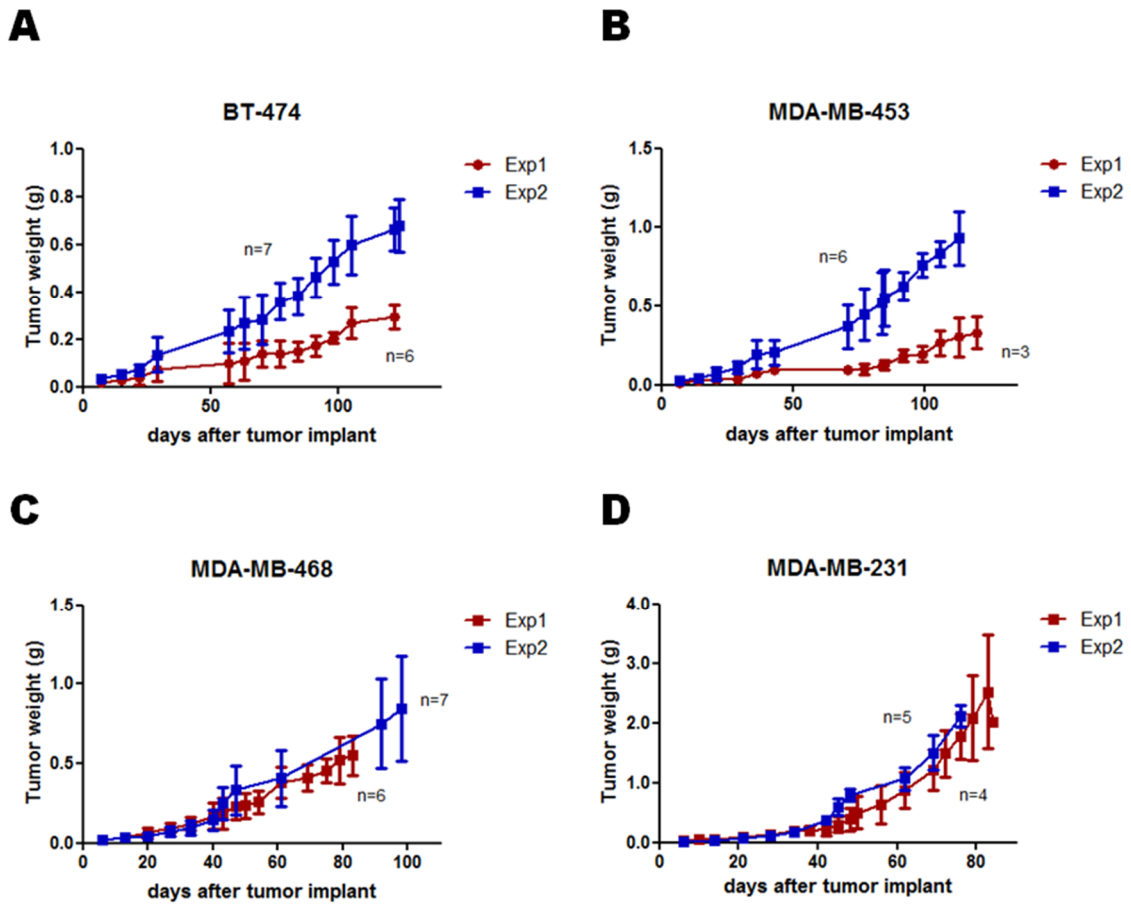


Figure 4.1.3. Tumor growth rate of BC cell lines in immunocompromised mice. Tumor growth curves represent the mean±SD mass of a single nodule for (**A**) BT-474 and (**B**) MDA-MB-453, and the mean±SD of the sum of the masses of two nodules for (**C**) MDA-MB-468 and (**D**) MDA-MB-231 models.

Overall, in both experiments, the number of single CTCs and CTM per milliliter of blood at sacrifice was higher in the MDA-MB-231 compared to the other CTC-models, in keeping with the aggressiveness and high proliferation rate of these cells, although high variability in CTC load was observed for both single CTC or CTM subpopulations (Figure 4.1.4, Panels A and B, and Table 4.1.1). Moreover, CTM frequency in CTC-positive samples was approximately 3-fold lower in less aggressive BC models compared to MDA-MB-231 (7 out of 24 *vs.* 6 out of 7 CTM-positive out of the total number of assessable cases, respectively). Interestingly, CTCs with different morphology within the same sample could be observed in all models, thus suggesting the existence of different subsets within the single CTC and CTC

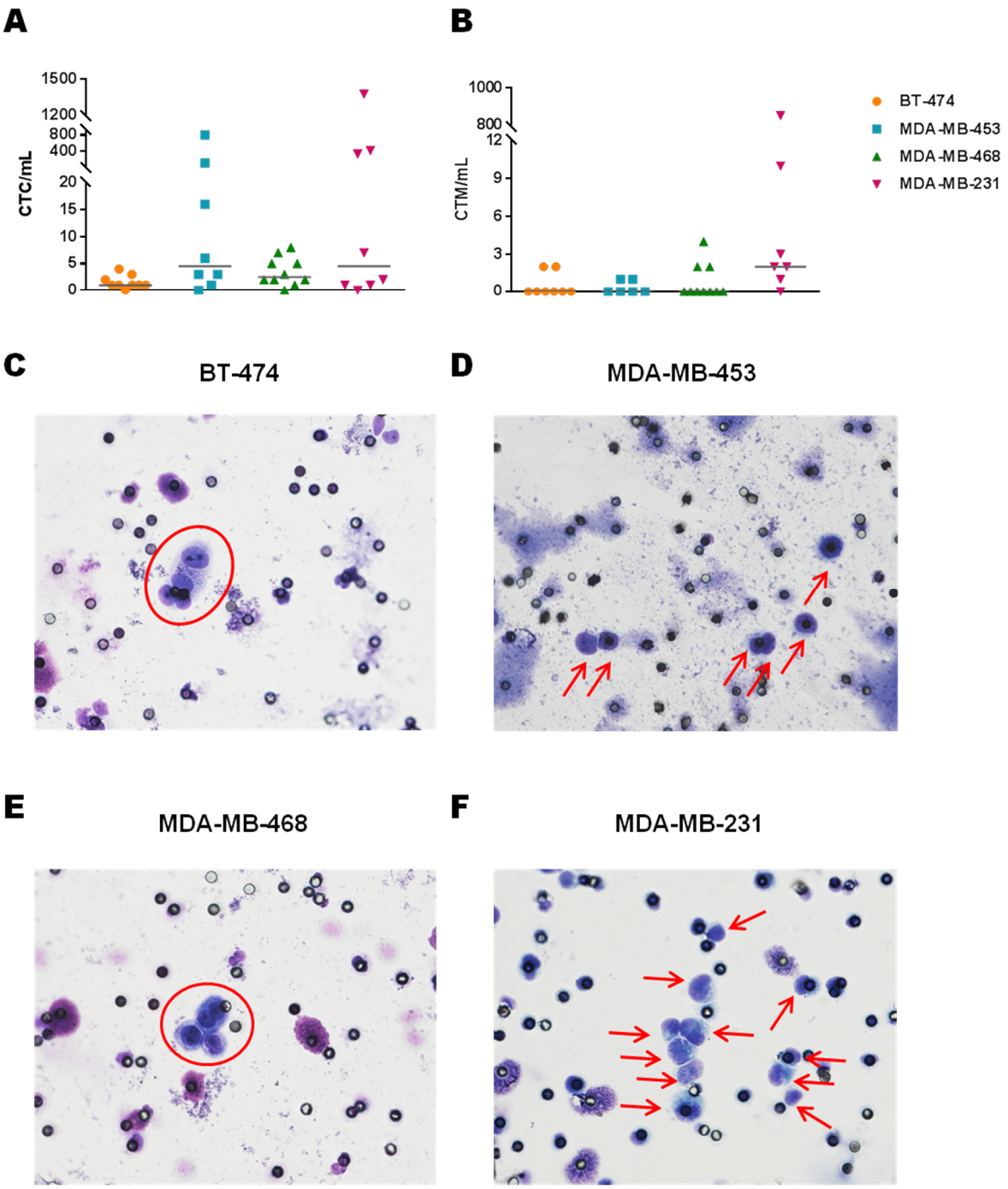


Figure 4.1.4. CTC and CTM load in BC xenograft models. Box plots represent the distribution of (**A**) CTCs and (**B**) CTM density, and images (400X total magnification) represent CTCs and CTM from cytological blood samples of (**C**) BT-474, (**D**) MDA-MB-453, (**E**) MDA-MB-468 and (**F**) MDA-MB-231 xenograft models. Red circle and arrows indicate single CTCs and CTC clusters/CTM, respectively.

cluster subpopulations (Figure 4.1.4, Panels C-F): image in Panel C shows a group of 4 CTCs from the BT-474 model forming a cluster intermingled with some platelets; the 6 single CTCs from the MDA-MB-453 model (Panel D) were different in size; a cluster of 4 CTCs and

surrounding platelets from the MDA-MB-468 model is depicted in Panel E, showing that heterogeneity in size can be observed also within tumor cells forming a CTM; the degree of pleomorphism is particularly clear in the MDA-MB-231 model (Panel F), where 11 single CTCs are represented, each with a more or less irregularly outlined nucleus of various size and shape, in addition to the heterogeneity in the whole cell morphology.

The metastatic potential of BC cell lines (Table 4.1.2 and Figure 4.1.5) was preliminary assessed by macroscopic inspection and IHC analysis on a series of non-adjacent FFPE sections (series of 4 consecutive stained and 8 consecutive unstained sections) obtained from different organs in Exp1. This preliminary experiment allowed identifying organs with high frequency of metastases, as lung, lymph-nodes and ovaries, and those without metastatic involvement, as liver, brain and spleen. In the second experiment, systematic IHC analysis was focused on a series of 24 or 48 non-adjacent FFPE sections (series of 8 consecutive stained and 8 consecutive unstained sections) from lung, lymph-nodes and ovary samples. The presence of metastatic foci at lung was observed in roughly all of the examined FFPE sections and in the majority of animals per experimental model. Ovarian metastases were detectable in MDA-MB-468 and MDA-MB-453, but not BT-474 models, whereas LNs were not detectable at macroscopic level in BT-474 and MDA-MB-453. On the contrary, lymph-nodal involvement was already visible macroscopically and at high frequency during sacrifice and organs explant in MDA-MB-468 and MDA-MB-231, as also confirmed by IHC analysis. Consistently with CTC load, all COX IV-positive lung sections from weakly aggressive models showed the presence of metastatic foci with smaller extent (single scattered cells or small foci of 3-30 cells) compared to MDA-MB-231 xenografts.

Taken together, these experiments provided information for the first time on the CTC load and the existence of subtypes of CTCs in different BC cell lines. Considering the limited CTC load in BT-474, MDA-MB-453 and MDA-MB-468 models, the molecular characterization of CTCs for this work of thesis (reported in “4.2. Molecular characterization of circulating tumor cells and solid tumor lesions in the MDA-MB-231 model”) was focused on the MDA-MB-231 model only.

Table 4.1.1. CTC load in BC xenograft models

CTC-model	CTC load					CTM load				
	Exp1		Exp2		Exp1 + Exp2	Exp1		Exp2		Exp1 + Exp2
	N*	median(IQR) or range CTC/mL	N*	median(IQR) or range CTC/mL	overall mean±SD CTC/mL	N*	median(IQR) or range CTM/mL	N*	median(IQR) or range CTM/mL	overall mean±SD CTM/mL
BT-474	2	0-2	7	1(1-2)	1.56±1.24	2	0-2	6	0	0.50±0.93
MDA-MB-453	3	0-111	5	6(3-16)	117.50±278.31	1	0	5	0(0-1)	0.33±0.52
MDA-MB-468	3	2-5	7	2(1-6)	3.50±2.64	3	0-4	7	0(0-1)	0.80±1.40
MDA-MB-231	5	7(1-333)	3	1-417	267.00±478.86	4	1-853	3	0-10	124.43±321.29

* number of animals assessable for CTC and/or CTM count according to the quality of the cytological sample

Table 4.1.2. Metastasis sites and frequencies in breast cancer xenograft models

CTC-model	Exp1									Exp2								
	N	Lymph-node		Lungs		Ovary 1		Ovary 2		N	Lymph-node		Lungs		Ovary 1		Ovary 2	
		+ve cases	Positivity frequency (%)*	+ve cases	Positivity frequency (%)*	+ve cases	Positivity frequency (%)*	+ve cases	Positivity frequency (%)*		+ve cases	Positivity frequency (%)*	+ve cases	Positivity frequency (%)*	+ve cases	Positivity frequency (%)*	+ve cases	Positivity frequency (%)*
BT-474	6	-§	-	6/6	25-100	0/5	-	0/5	-	7	-§	-	7/7	100	0/7	-	0/7	-
MDA-MB-453	3	-§	-	2/3	80-100	0/3	-	0/3	-	6	-§	-	6/6	75-100	6/6	75-100	5/5‡	43-100
MDA-MB-468	6	2/2	36-100	6/6	8-100	1/4	21	1/4	21	7	7/7	100	7/7	100	2/7	23-100	1/7	90
MDA-MB-231	4	4/4	100	4/4	100	-§	-	-§	-	5	5	100	5/5	100	-	-	-	-

* range of positivity frequencies (number of metastasis +ve sections: from 10 to 30 sections in Exp1, and from 24 to 48 sections in Exp2)

§ not available or not detectable at macroscopic level

‡ 1 out of 2 ovaries was not assessable

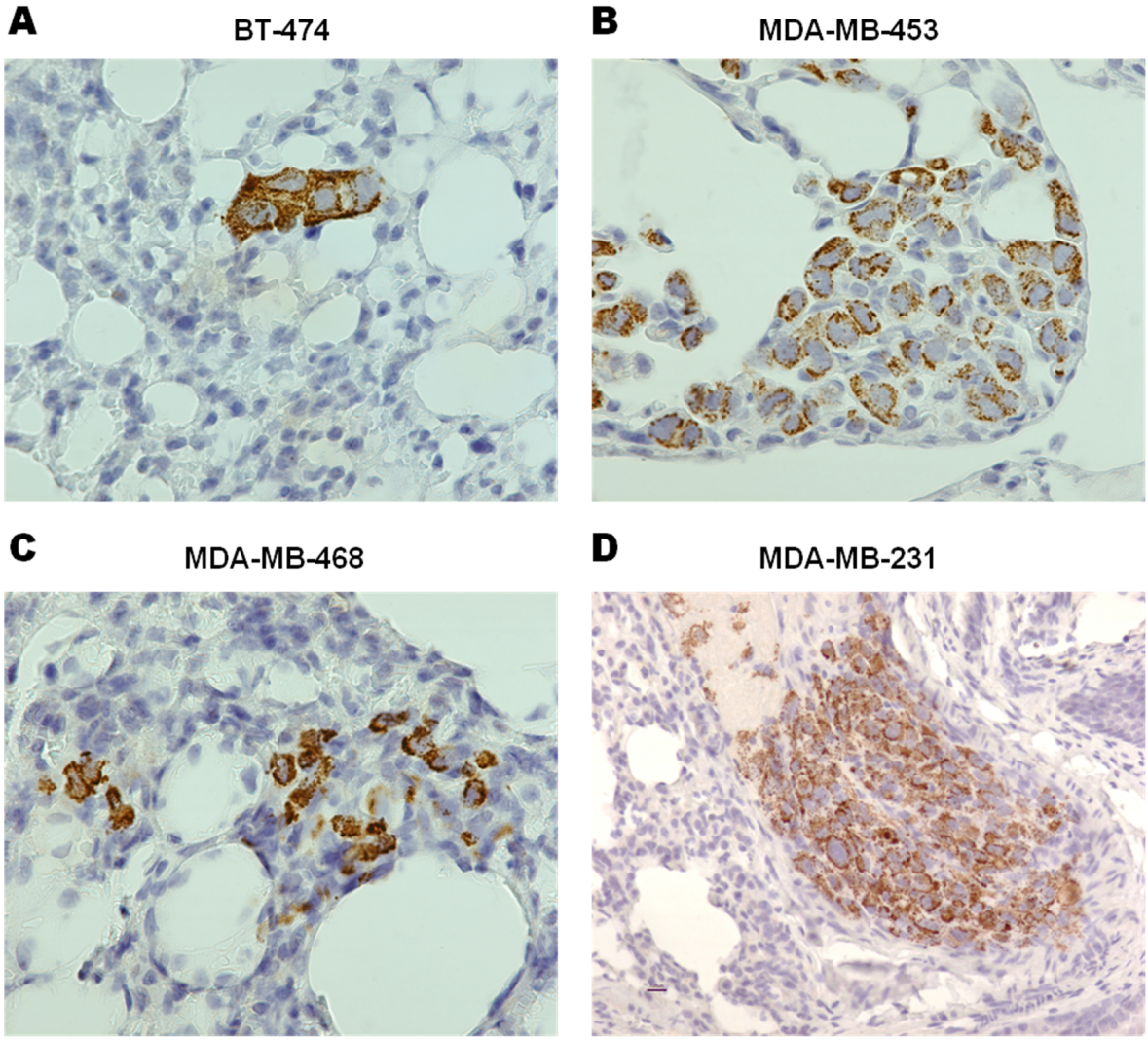


Figure 4.1.5. Metastatic burden of breast cancer cell lines after orthotopic injection in mice. Images represent COX IV stained metastatic cells in a FFPE section of lung samples from (**A**) BT-474, (**B**) MDA-MB-453, (**C**) MDA-MB-468 (400X total magnification, oil immersion) and (**D**) MDA-MB-231 (200X total magnification) xenograft models.

4.1.3. Longitudinal analysis of CTC load and metastatic burden in the MDA-MB-231 xenograft model

The dynamics of dissemination in the MDA-MB-231 model was investigated by a time-course experiment, where CTC load and metastatic burden were measured at different time points after tumor cell injection. This experiment was designed with the final aim to isolate early and late CTCs, *i.e.* released before and after the completion of the metastatic

cascade, and to compare their molecular profile, on the hypothesis that temporal heterogeneity could be modeled according to the phase of dissemination, an early dormant followed by an active and proliferating one.

Overall, primary tumor weight (Figure 4.1.6, Panel A), CTC load (Figure 4.1.6, Panel B) and CTM load (Figure 4.1.6, Panel C) showed a step-wise increase at each experimental time point, although the experimental variability for each biological feature was lower in the early phases (day 35 and day 50) compared to the late phases (day 65 and day 80) of the time-course. Following a similar trend, the frequency of metastasis-positive cases assessed in lymph-node (axillary, inguinal, subclavean or peritoneal) and lung samples increased during time, showing the highest positivity frequency during the last two time points, although metastases at lymph-nodes, differently than lungs, were detectable since the earliest phase of tumor progression in the majority of assessable cases (Figure 4.1.6, Panel D). At day 35 CTCs were found in 1 out of 5 assessable cases (2 CTCs, Figure 4.1.6, Panel B), consistently with the detection of few metastatic cells at lung in the same animal (Figure 4.1.6, Panel D). At day 50 concordance between CTC-positive and matched metastases-positive lung samples was low as 1 out of 3 CTC-positive cases only had metastases. On the contrary, 5 out of 6 cases at day 65 and 5 out of 5 cases at day 80 had both CTCs and pulmonary metastases.

Longitudinal analysis of dissemination in the MDA-MB-231 CTC-model allowed the identification of a temporal window, from day 35 to day 50 after tumor implant, during which CTCs but not metastases at lung can be found, although complete inspection of lung samples is required to confirm this result. However, at the same time, the outcome of the experiment revealed that a complete molecular characterization of the CTC-model during each step of the tumor progression is not feasible with available techniques due to the limited number of CTCs isolated in the early phases, and that a higher number of animals should be sacrificed to obtain representative CTC samples. For these reasons, the molecular characterization of the CTC-model, reported in the following chapter, was performed at the endpoint (around day 80) of the experiment.

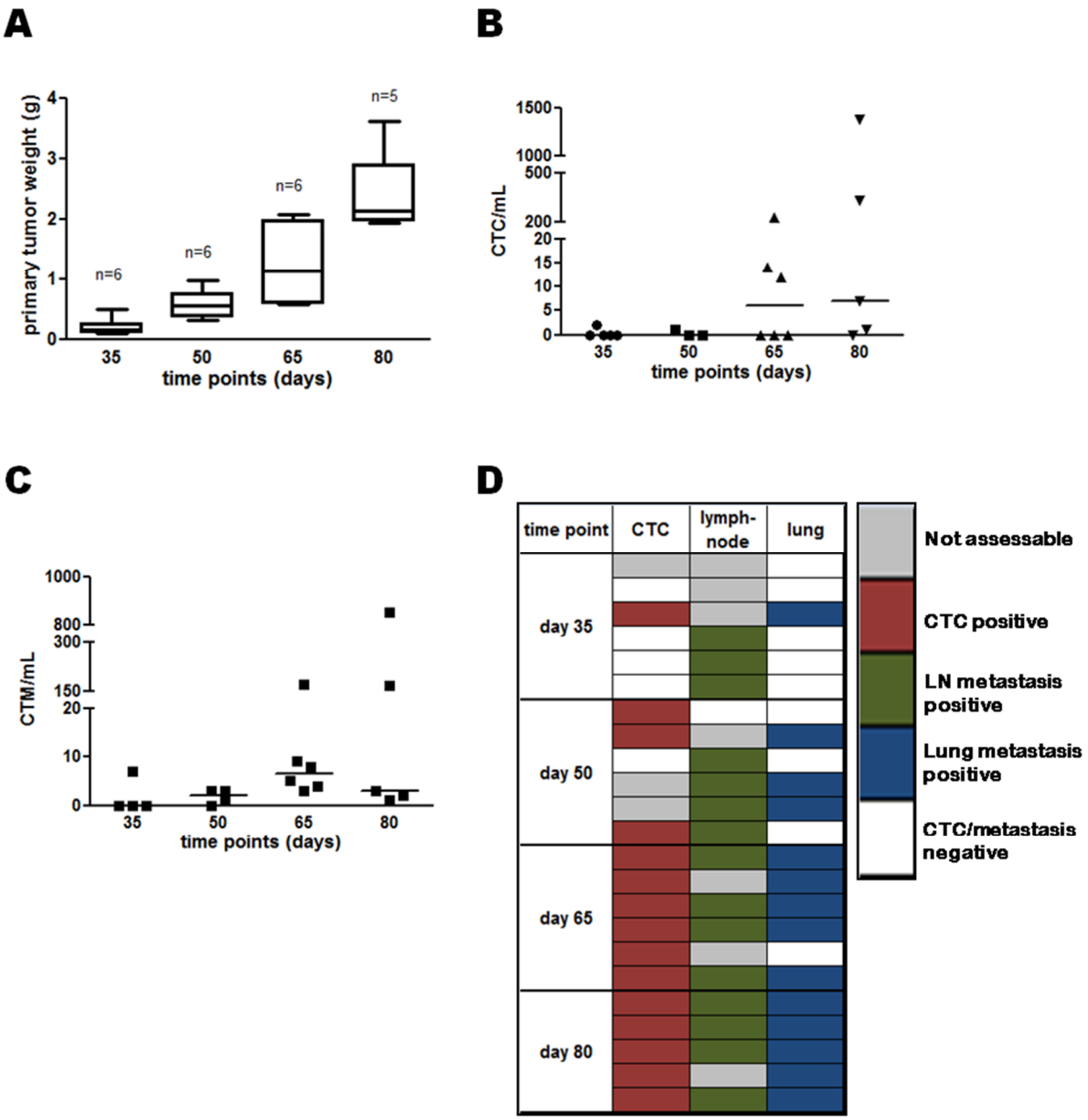


Figure 4.1.6. Longitudinal analysis of dissemination in the MDA-MB-231 xenograft model. Box/dot plots represent the distribution of (**A**) the total tumor mass (from two nodules), and (**B**) CTCs and (**C**) CTM at different experimental time points. Scheme (**D**) represents the frequency of CTC/CTM or LN or lung metastasis positive samples for each animal, according to the experimental time point.

On the hypothesis that CTCs isolated during the late phase of metastasis could derive from both the PT and metastases at lung, MDA-MB-231 cells were also intravenously injected in 5 animals and their presence in blood was monitored during time. Blood samples collected at sacrifice 1 hour after injection from one animal injected with 1 million and one animal injected with 2 million cells contained 1 and 7 CTCs per milliliter, respectively, thus indicating that the vast majority of cells had reached peripheral districts. The remaining

three animals, two injected with 1 million and one injected with 2 million cells, were sacrificed after 90 days and were all CTC-positive and lung metastasis-positive. CTM numbers were detected in all cases and ranged from 1 to 31 per milliliter, while CTCs (about 280) were found in one animal only, injected with 1 million cells. Lymph-nodes, ovaries and spleen were all negative for metastases by macroscopic examination and IHC of 4 FFPE consecutive sections.

These results show that our model does not completely recapitulate the tumor progression steps observed in the clinical setting as MDA-MB-231 cells, once in the bloodstream, colonize the lungs rapidly. Moreover, their behavior might have implications for molecular profiling studies since lung metastases represent another source of CTCs in addition to the primary site.

4.2. Molecular characterization of circulating tumor cells and solid tumor lesions in the MDA-MB-231 model

4.2.1. Reliable PCR-based quantification and gene expression microarray-based analysis of tumor cells in biological samples derived from xenograft models are feasible

A series of preliminary tests was run before performing gene expression profile (GEP) experiments with the MDA-MB-231 xenograft model in order to address technical issues related to the sensitivity and specificity of the assays applied for detection, quantification and characterization of tumor cells in biological samples from xenograft mouse models.

The first limiting step when selecting samples for GEP of CTC-models is the availability of a sufficiently representative biological sample and numbers of tumor cells superior to the technical detection limit of the platform applied for molecular analysis. In order to detect CTCs, which notoriously are rare events, and to confidently discriminate between CTC-positive and CTC-negative samples by nucleic acid analysis, tumor cell isolation was performed using immunomagnetic beads coupled to a size-based method (AdnaTest and ScreenCell® Molecular Biology kits). CTC detection was performed using a qPCR-based approach, taking advantage of the species-specificity of commercially available TaqMan gene expression assays. Searching for assays which theoretically should not detect murine transcripts and designed for a gene expressed at equal levels in tumor cells, a particular TaqMan assay for human *ACTB* constitutive gene was selected and its species-specificity verified in a qPCR test with universal murine RNA reference (UMR). Gene encoding for murine β-actin was never detected, either when amplifying 10 or 50 ng of cDNA (data not shown).

As second phase, the feasibility of the whole approach for CTC detection was assessed using cDNA of blood samples from three MDA-MB-231 xenograft models processed using a 3-step approach, consisting of incubation with immunomagnetic beads

provided with 1) AdnaTest EMT-1 Select kit, as first step, and with 2) from AdnaTest EMT-2 Select kit, as second step, followed 3) isolation size-based with the ScreenCell® Molecular Biology kit as third step. A total of 8 samples were obtained (1 case processed with AdnaTest EMT-2 and ScreenCell only). Samples were considered as positive for the presence of tumor cells when C_t values were lower than 40 after 40 cycles of amplification. Quality control test for negative samples was performed using a TaqMan assay specific for murine *Gapdh*, previously verified to be specific and not able to detect human transcripts in a qPCR test with universal human RNA reference (UHR), in order to exclude possible technical failures during the tumor cell isolation procedure, especially for CTC samples. Overall, irrespective of the approach used for CTC isolation, C_t values for human *ACTB* ranged from 16.78 to 39.35 (median 27.04) in 7 samples, whereas 1 out of 8 samples only did not reach the technical threshold value. The CTC-negative sample was submitted to quality control test using a TaqMan assay for murine endogenous control gene: C_t value for murine *Gapdh* was 22.71, thus indicating that the result was not due to a technical failure and confirming the negative CTC status. Therefore, the validity of the method for MDA-MB-231 CTC isolation (*i.e.*, by a positive selection step followed by a size-based isolation step) and detection (*i.e.*, by qPCR) was in this manner verified.

Considering its high specificity, the molecular approach designed for CTC detection was supposed to be applicable to the estimation of the number of tumor cells not only contained in blood samples but also embedded in other tissues and organs, as tumor nodules and metastasis positive organs collected from mice at their sacrifice. This kind of test, other than representing an alternative approach for tumor cell quantification *per se* (beside direct visualization of tumor cells by image-based techniques), is fundamental to reach similar experimental conditions during transcripts amplification and hybridization in GEP experiments with high-density oligonucleotide arrays, as it allows input of equal human RNA quantities on the microarray platform, information that cannot be obtained from mixed murine and human RNA samples by using direct RNA quantification methods. In order to indirectly calculate the number of tumor cells a standard curve was generated using different

amounts of MDA-MB-231 cells (2-5-10-20-40-100-200-1,000-5,000-10,000-50,000-100,000) and correlating them to C_t values obtained by qPCR for human *ACTB* (Figure 4.2.1). Observed and expected C_t values showed high correlation (Pearson's correlation r = 0.9960), indicating that the approach allows reliable quantification of the total number of tumor cells in samples containing both human and murine RNA, thanks to the good performance of the assay and to homogenous expression of *ACTB* in MDA-MB-231 cells.

Enrichment methods that are able to completely exclude hematopoietic or other non-tumor cells do not exist nowadays. Nonetheless, CTCs can be detected in whole blood samples even when present at low densities. Moreover, the sensitivity and the specificity of the methodological approach used for CTC analysis should be preliminarily evaluated, in order to ensure adequate sampling and reliable characterization. In our laboratory, spiking experiments demonstrated that the WG DASL HT assay (Illumina) has high sensitivity, since it allows obtaining reliable GEP from a minimum of 25 tumor cells spiked in a 5-mL sample of whole blood of healthy donor and captured with the AdnaTest EMT-1 Select kit (Fina et al., 2015a), which consists of immunomagnetic beads for enrichment of CTCs and a wash buffer that reduces by 10-fold contamination from non-target cells.

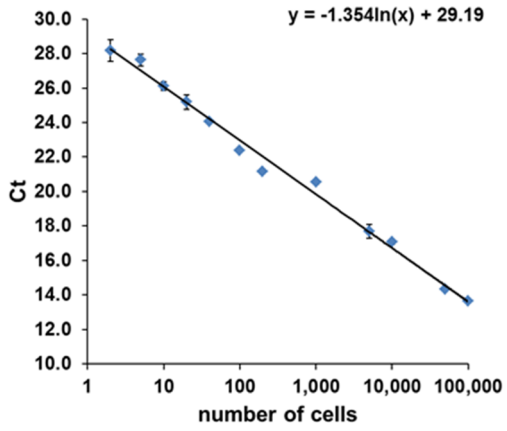


Figure 4.2.1. Method for estimation of the number of tumor cells in samples from MDA-MB-231 xenograft models by relative quantification of human *ACTB*. Standard curve was constructed by interpolating different numbers of MDA-MB-231 cells with corresponding Ct values obtained by qPCR for h*ACTB*. Mean±SD C_t values of three independent experiments are reported as a function of cell number.

Although the approach developed for CTC analysis in human samples can be applied also in the experimental setting, it should be considered that there is no possibility to minimize the contamination of stromal cells in samples of solid tumor lesions obtained from xenograft models, unless microdissection or tumor cell disaggregation from fresh or FFPE samples are used. In such a case, however, the assessment of the species-specificity of molecular probes can represent the solution. Therefore, a specificity test was performed to assess if probes of the DASL assay, designed for human transcriptome, also recognize transcripts of murine origin. To this aim, considering that the amount of murine mRNA in samples from xenograft models might be variable according to the site and extension of solid lesion (PT or micro- and macro-metastases) and that contaminating murine mRNA could affect the reliability of the GEP at different extent, a test was designed in an attempt to mimic the expected composition of the biological sample. To this aim, pure universal RNA reference from human (UHR) or murine (UMR) tissues and mixtures of them at 1:3, 1:1 and 3:1 proportions to a maximum of 50 ng of total RNA were processed with the DASL assay. All samples containing UHR had roughly 70% detection rate, independently from the contribution of UHR, whereas pure UMR samples had low detection rates (10 and 30%) and signal intensities 3-fold lower than those obtained from the other samples (Figure 4.2.2, Panel A). Consistently, reciprocal correlation analysis showed that pure UMR and samples containing UHR form two distinct clusters of highly correlated samples (Figure 4.2.2, Panel B), thus confirming that the GEP obtained from xenograft samples cannot be affected by the presence of murine mRNA.

This test revealed that the DASL assay is sufficiently specific for human transcriptome to ensure reliable molecular characterization of CTCs and solid lesions from xenograft models. Nevertheless, considering a detection rate of 10% and 30% in the two pure UMR samples even if at low intensities, qPCR tests with specific TaqMan assays might be required to validate the expression of some genes.

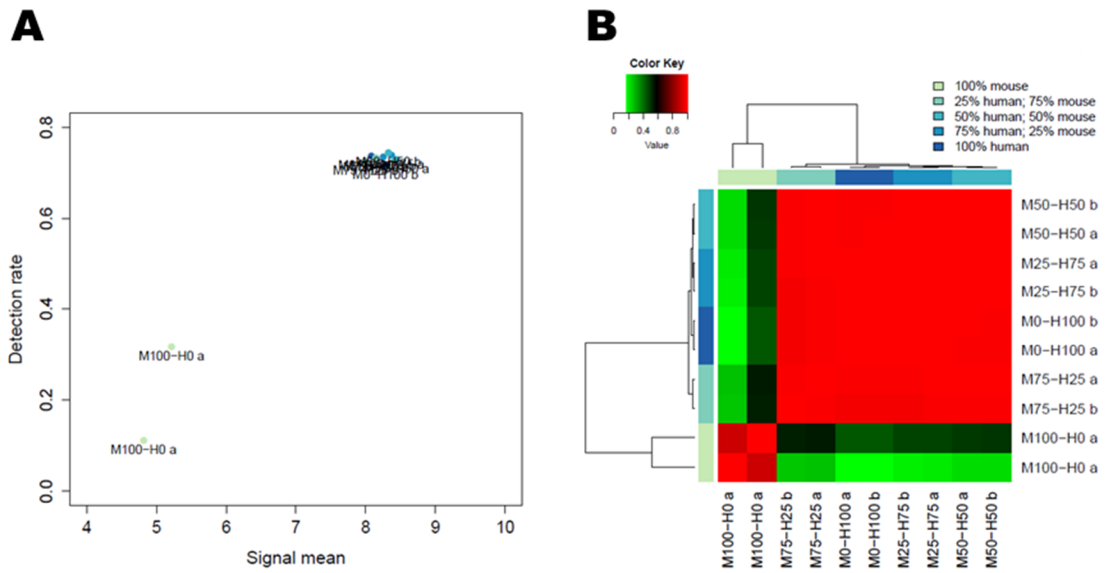


Figure 4.2.2. Specificity of the WG DASL HT assay platform for molecular analysis of human samples. (**A**) Scatter plot represents mean signal intensities and detection rates. (**B**) Heat map represents pair-wise correlations between gene expression profiles for duplicate samples ("a" and "b") containing different proportions of UHR ("H") and UMR ("M") (numbers indicate the percentage).

4.2.2. Circulating tumor cells have a distinct transcriptome profile compared to solid primary and secondary lesions

On the hypothesis that comparing the molecular profile of CTCs with those of the PT and metastases at lungs (LUNG) and lymph-nodes (LN) could allow obtaining new information on the pathways required for cell dissemination and defining a set of biomarkers with clinical relevance in breast cancer, the MDA-MB-231 xenograft model was extensively characterized using the WG DASL HT assay after two independent experiments (hereafter referred to as "GEP1" and "GEP2"), performed in 2013 and 2015. The GEP of CTC was compared with those obtained from LN sections and sections from peripheral and central sites (in the attempt to obtain sufficiently representative samples and do not disregard possible differences due to heterogeneity) of PT and LUNG. In addition, the analysis was extended to the characterization of the parental MDA-MB-231 cells, collected after standard *in vitro* culture conditions, as such information was considered instrumental to select candidate genes for preliminary functional studies by *in vitro* assays.

Ten animals per experiment were injected with MDA-MB-231 cells and CTCs were indirectly quantified by qPCR on the basis of the standard curve previously created for this study model (results are reported in Table 4.2.1). Biological samples for GEP analysis were selected from three mice according to the total number of yielded CTCs, in order to profile the highest number of tumor cells available. Subsequently, matched samples of PT, LN and lungs were processed 1) for preliminary assessment of tumor cell presence by IHC analysis on OCT-frozen sections and 2) for indirect quantification of tumor cells in consecutive sections of each positive sample. Considering the maximum amount of detected CTCs, volumes of RNA solution corresponding to 5,000 tumor cells according to qPCR analysis for human *ACTB* were calculated for each sample and used as RNA input. On the average, 50 ng of total RNA were used as input, since this amount yields high gene detection and reliable profiling analysis with the DASL assay (Fina et al., 2015a).

Table 4.2.1. CTC load estimated by indirect quantification

GEP1 experiment			GEP2 experiment*	
Animal ID	CTC total number	DTC total number	Animal ID	CTC total number
63X	0	NA	**147X**	9,137
65X	3	0	148X	25
66X	30	0	149X	268
67X	0	NA	150X	821
68X	21,811	NA	151X	565
69X	36,092	17,884	**152X**	15,057
70X	34,431	54	154X	172
76X	1,714	NA	**155X**	62,200
77X	48	NA	156X	513
78X	172	262	157X	5,442

* DTC isolation not performed

NA: not available

Preliminary assessment of signal intensities obtained after raw data acquisition (Figure 4.2.3) confirmed that low cross-reactivity of the assay with murine RNA occurred in both experiments, since in UMR control samples median signal intensities approximated the lower limit of detection (corresponding to 4), and the number of detected genes was negligible compared to those obtained in experimental samples.

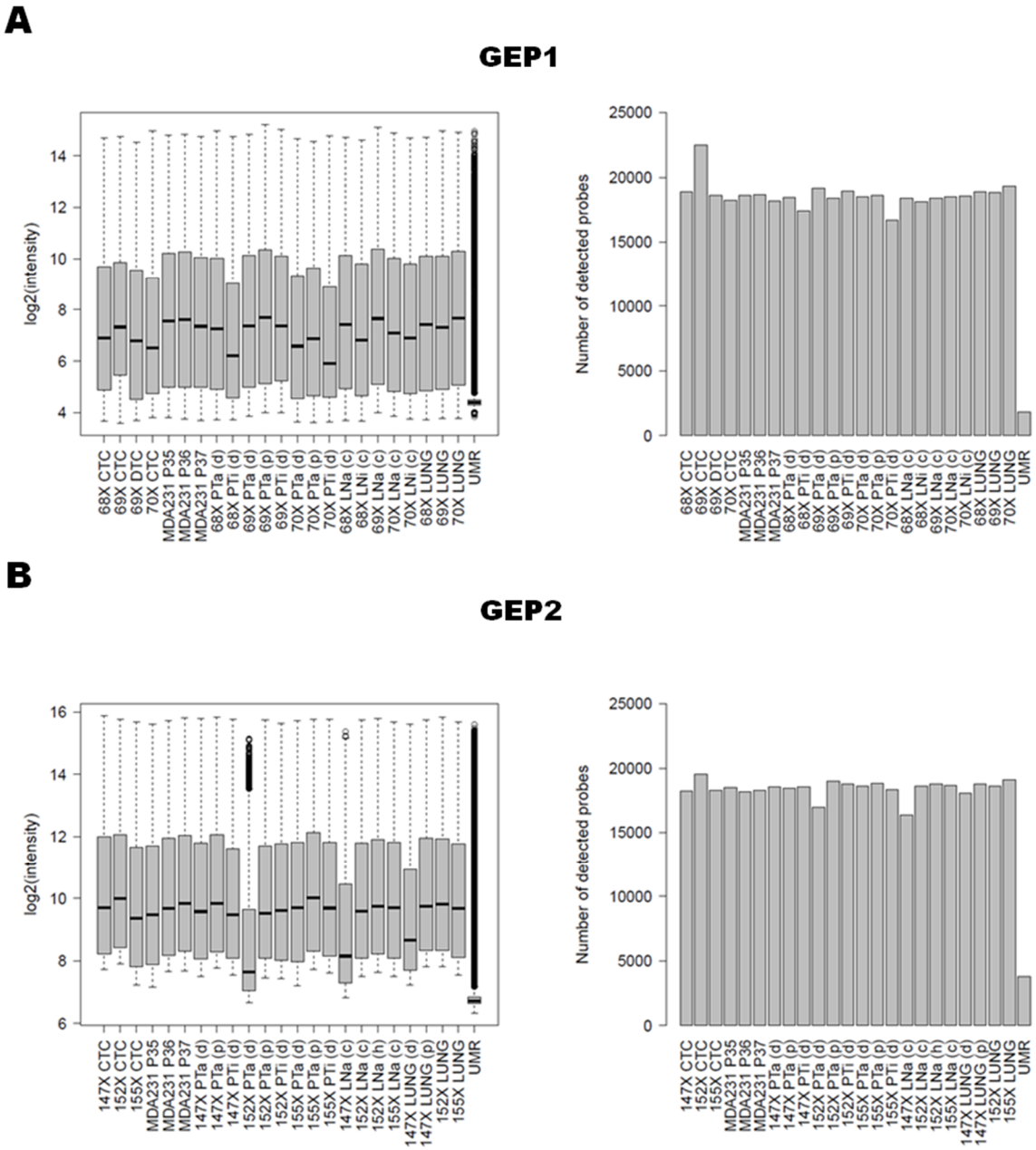


Figure 4.2.3. Raw signal intensities and detection rates for gene expression profiles of MDA-MB-231 CTC-model. Box plots (*left*) represent the distribution of raw signals and bar charts (*right*) display the number of detected genes for each sample in (**A**) GEP1 and (**B**) GEP2 experiments. “CTC”, circulating tumor cells; “DTC”, tumor cells disseminated to bone marrow; “PT”, primary tumor; “LN”, lymph nodal metastases; “LUNG”, pulmonary metastases; “a”, axillary; “i”, inguinal; “(c)”, controlateral to PT; “(h)”, homolateral to PT; “(d)”, distal; “(p)”, proximal.

The analysis of reciprocal correlations among all samples (excluding UMR) revealed that one sample in GEP1 (corresponding to a peripheral section of inguinal PT obtained from animal 70X), and three samples in GEP2 (one corresponding to a peripheral section of axillary PT obtained from animal 152X, and the other two corresponding to a section of an axillary LN contralateral to the tumor and to a peripheral section of lungs of the same animal, ID 147X), were scarcely correlated with the other samples (Figure 4.2.4). Since distributions of raw intensities in the aforementioned samples were overall lower (Figure 4.2.3), technical

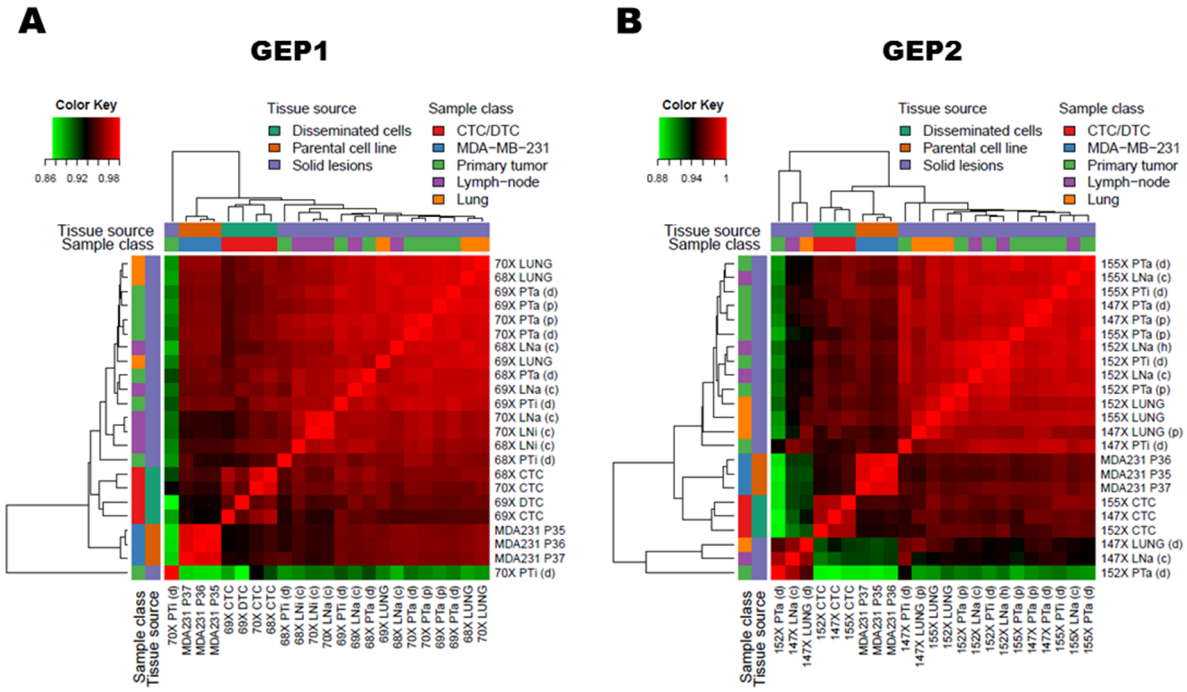


Figure 4.2.4. Reciprocal correlations between GEP raw data of samples derived from the MDA-MB-231 CTC-model. Heat maps represent pair-wise correlations of raw gene expression data from (**A**) GEP1 and (**B**) GEP2 experiments. “CTC”, circulating tumor cells; “DTC”, tumor cells disseminated to bone marrow; “PT”, primary tumor; “LN”, lymph nodal metastases; “LUNG”, pulmonary metastases; “a”, axillary; “i”, inguinal; “(c)”, controlateral to PT; “(h)”, homolateral to PT; “(d)”, distal; “(p)”, proximal.

failure or suboptimal performance of the assay was considered a more plausible cause of their deviance rather than biological variability or heterogeneity among different sites of the PT. Importantly, within the main cluster, three sub-clusters were identifiable in both experiments, as represented in the heat map in Figure 4.2.4: 1) one including samples

obtained from captured cells (*i.e.*, CTC and DTC), 2) one corresponding to cultured cells (*i.e.*, MDA-MB-231 parental cell line), and 3) one including all the remaining samples from solid lesions, excluded those displaying low intensities. On the basis of these considerations, deviating samples were excluded. Therefore, 22 and 20 samples respectively from GEP1 and GEP2 were included in subsequent analyses.

Signal intensities of samples which passed the above described quality control test were again analyzed after data normalization. Analysis of distribution of normalized (RSN) data revealed homogeneous signal intensities among samples in both experiments (Figure 4.2.5), again confirming the technical validity of the indirect quantification procedure to ensure comparable RNA inputs.

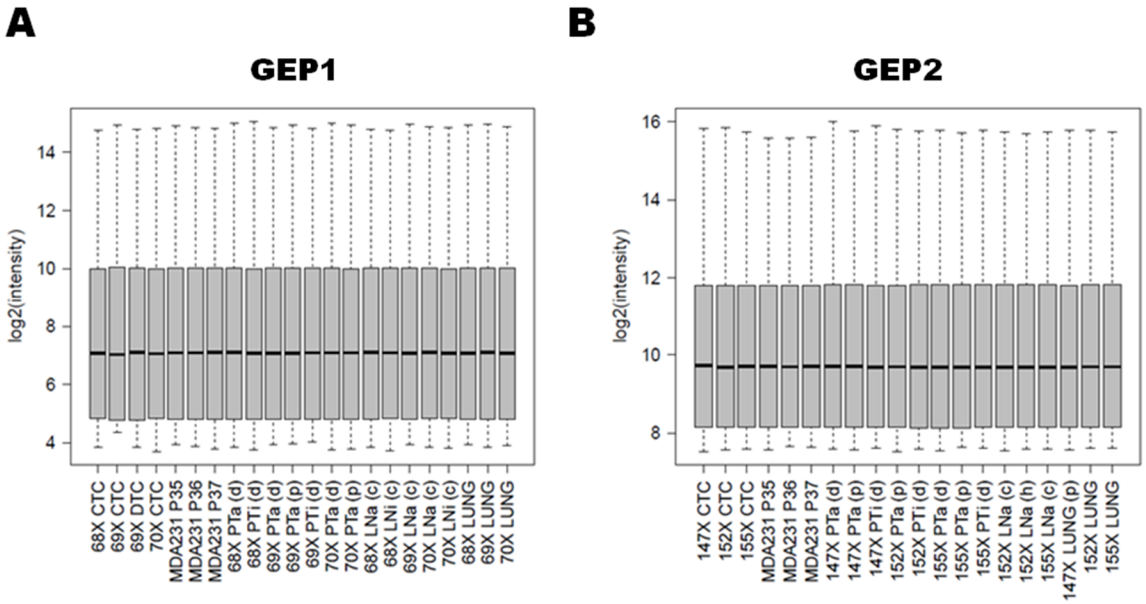


Figure 4.2.5. Normalized signal intensities of MDA-MB-231 CTC-model profiled samples. Box plots represent distribution of RSN intensities in (**A**) GEP1 and (**B**) GEP2 experiments. “CTC”, circulating tumor cells; “DTC”, tumor cells disseminated to bone marrow; “PT”, primary tumor; “LN”, lymph nodal metastases; “LUNG”, pulmonary metastases; “a”, axillary; “i”, inguinal; “(c)”, controlateral to PT; “(h)”, homolateral to PT; “(d)”, distal; “(p)”, proximal.

Again, three clusters emerged from the correlation analysis of normalized data, including disseminated tumor cells, solid tumor lesions or the parental cell line (Figure 4.2.6). Importantly, all the sections obtained from solid lesions were highly correlated and no clear

sub-clusters could be identified either according to the sample class (PT, LUNG or LN) or the site of the lesion (peripheral or central, as assessed for PT and LUNG samples), thus indicating that tumor lesions share similar expression patterns which are distinct from those peculiar of disseminated cells, and even suggesting that after hematogenous dissemination, tumor cells colonizing distant organs tend to recapitulate the molecular features of the PT.

In order to identify genes that allow distinguishing CTCs from the other lesions and from the parental cell line, unsupervised hierarchical clustering was performed using the most variable genes, *i.e.* those with IQR intensities >95th percentile (corresponding to 209 genes in both experiments). Also in this case, samples segregated in three clusters,

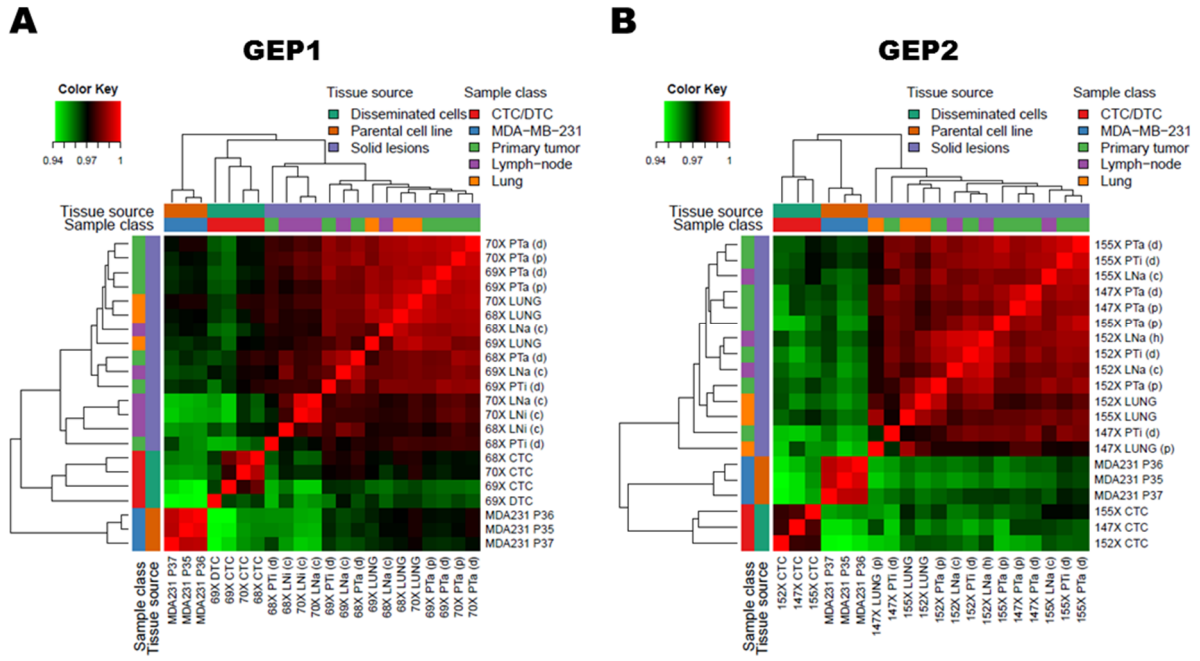


Figure 4.2.6. Reciprocal correlations between GEP RSN data of samples derived from the MDA-MB-231 CTC-model. Heat maps represent pair-wise correlations using RSN expression data from (**A**) GEP1 and (**B**) GEP2 experiments. CTC", circulating tumor cells; "DTC", tumor cells disseminated to bone marrow; "PT", primary tumor; "LN", lymph nodal metastases; "LUNG", pulmonary metastases; "a", axillary; "i", inguinal; "(c)", controlateral to PT; "(h)", homolateral to PT; "(d)", distal; "(p)", proximal.

consisting of MDA-MB-231 cells, CTC samples or solid lesions, thus providing meaningful evidence that MDA-MB-231 cells, the tumor lesions they gave rise after orthotopic xenograft and the derived CTCs, despite the common origin, have distinct transcriptome profiles

(Figure 4.2.7). Considering the overall data, it was observed that MDA-MB-231 cells share about 50% of their highly expressed genes with CTCs, which in turn are characterized by a large portion (roughly 70% in GEP1, and ranging from 50 to 20%, according to the CTC sample, in GEP2) of genes with higher expression level compared to the parental cell line. Moreover, genes highly expressed in CTCs and those highly expressed in solid tumor lesions displayed slight overlap only, suggesting that molecular traits unique to CTCs exist. On the contrary, samples obtained from tissue sections shared the vast majority of genes and did not cluster according to the sample class (PT, LN and LUNG).

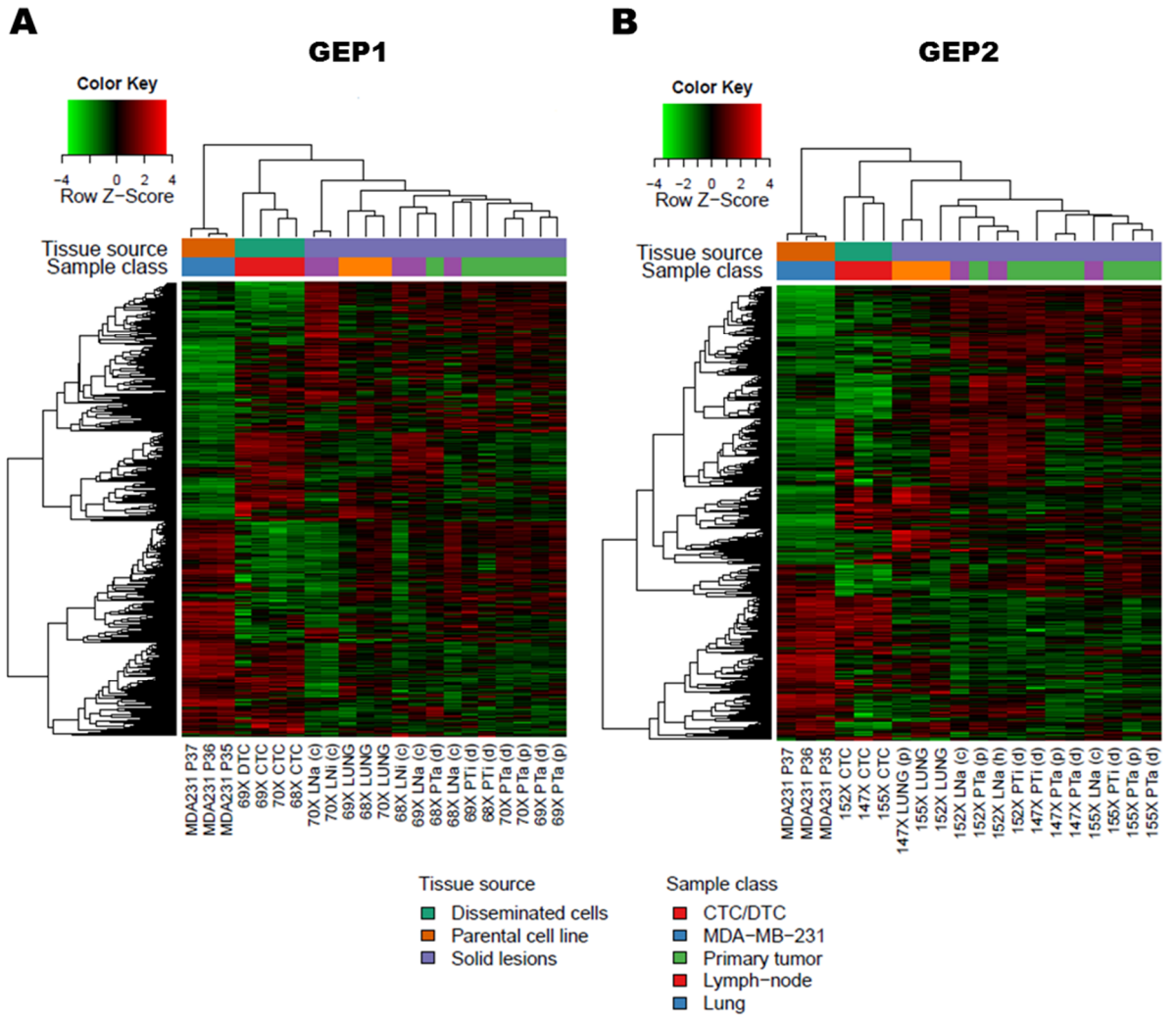


Figure 4.2.7. Hierarchical clustering analysis of GEPs of samples derived from the MDA-MB-231 CTC-model. Heat map representing the expression pattern of the most variable genes (IQR intensities >95th percentile) in (**A**) GEP1 and (**B**) GEP2 experiments.

To exclude the possibility that the biological variability among animals more than the variability among classes of samples could have confounded the interpretation of results, principal variant component analysis was performed considering the most plausible sources of experimental variability. The results of this explanatory analysis (Figure 4.2.8) confirmed that the major contribution to the different behavior of the three groups of samples was

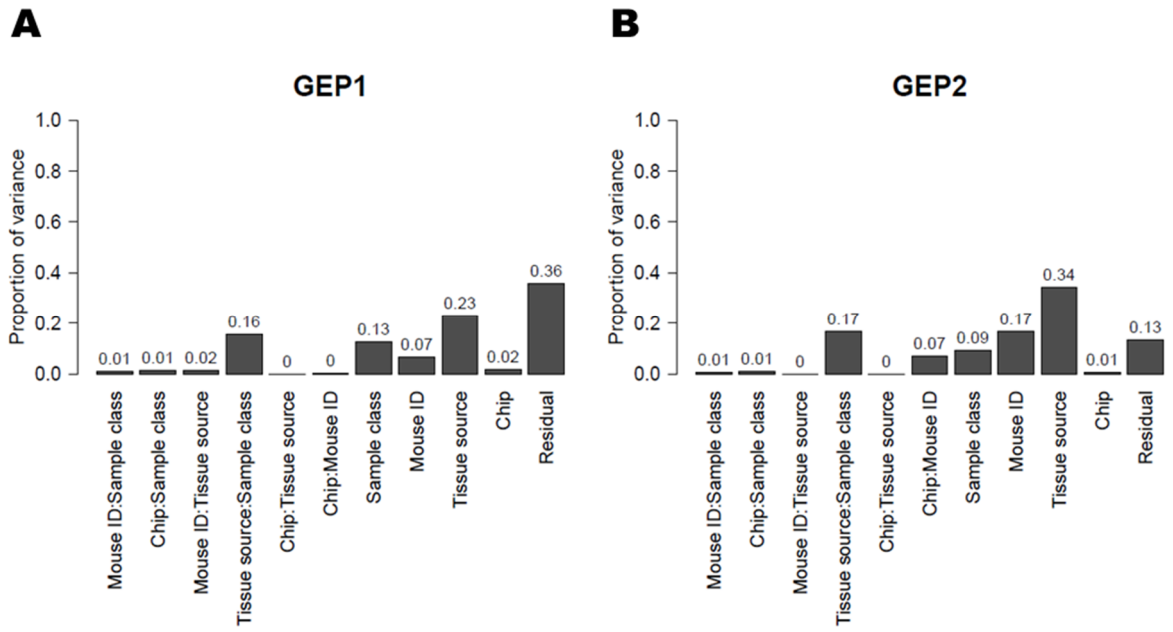


Figure 4.2.8. Sources of variability in GEP experiments with the MDA-MB-231 CTC-model. Bar charts represent the contribution of each investigated technical and biological factor to the global experimental variability in (**A**) GEP1 and (**B**) GEP2 experiments.

actually imputable to the tissue source (*i.e.*, disseminated cells, parental cells and solid lesions), accounting for 23% variability in GEP1 and for 34% in GEP2, rather than the inter-animal variability (7% and 17%) or the sample class (*i.e.*, CTC, PT, LUNG, LN and MDA-MB-231; 13% and 9%, respectively for GEP1 and GEP2). In addition, inter-assay variability was negligible (1% and 2%), similarly to data already reported in Fina et al., 2015a. These results indicate that the different transcriptome profile observed in CTCs compared to the other sample types reflects their distinct biological features rather than the result of an experimental artifact.

Correlations between log(base2) fold changes considering all detected genes in GEP1 and GEP2 experiments were acceptable as r = 0.66, 0.52, 0.56 and 0.75 (Pearson's correlation coefficient) were obtained respectively for CTC *vs.* PT or LN or LUNG or MDA-MB-231 pair-wise class comparisons. Of note, genes up-regulated and down-regulated in CTCs compared to all solid lesions and common to both experiments were 192 and 282, respectively.

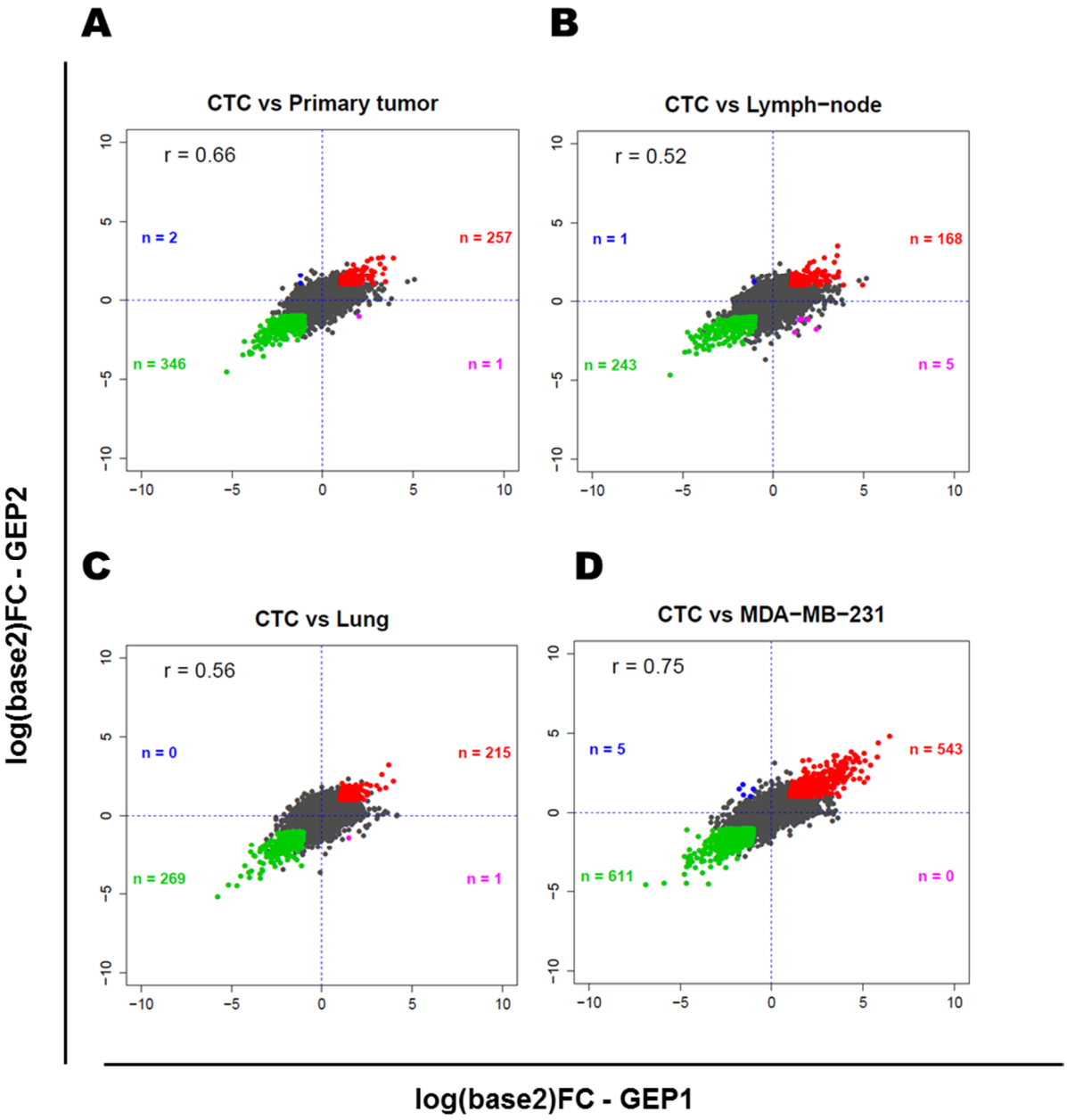


Figure 4.2.9. Correlation analysis between GEP1 and GEP2 experiments with MDA-MB-231 CTC-models. Scatter plots represent correlations among log(base2) fold changes considering all detected genes (indicated as dots) in GEP1 and GEP2 experiments, after (**A**) CTC *vs.* PT, (**B**) CTC *vs.* LN, (**C**) CTC *vs.* LUNG and (**D**) CTC *vs.* MDA-MB-231 class comparisons. Genes significantly

(Figure 4.2.9. continued) up-regulated or down-regulated (FC threshold = |2| and FDR <0.05) in both experiments are represented by red and green dots, respectively, and genes with discordant trend are represented by pink or blue dots if up-regulated in GEP1 and down-regulated in GEP2 or down-regulated in GEP1 and up-regulated in GEP2, respectively.

4.2.3. Genes peculiar to circulating tumor cells are mainly related to embryogenesis, development and chromatin remodeling

Extensive molecular characterization of the MDA-MB-231 CTC-model by comparing GEPs from PT, LUNG, LN, CTC and the parental cell line, offered the opportunity to better understand the molecular bases behind hematogenous dissemination and metastasis and generate hypotheses on the biological processes and pathways which might regulate the different phases of tumor progression, from tumor growth, to the early steps of dissemination, via lymphatic and hematogenous system, to colonization and secondary tumor establishment in distant organs.

Preliminary biological interpretation of GEP data was supported by GO analysis in all possible class comparisons, including CTC *vs.* solid lesions. A summary of GO terms mainly represented according to their frequency and significance is reported in Table 4.2.2. Genes significantly differentially expressed (FC threshold = |2| and FDR <0.05) in CTC compared to all solid lesions in GEP1 and GEP2 were 474 (192 up-regulated).

Genes up-regulated in CTCs compared to all solid lesions were enriched in GO terms related to embryogenesis, stem cell proliferation, development and morphogenesis of various tissues and organs, especially neural, renal and vascular system, suggesting pronounced plasticity in this CTC-model. Moreover, also GO terms related to response to physical, chemical and biological external stimuli, including response to pathogens, were highly represented (Figure 4.2.10, Panel A). Interestingly, GO terms related to metabolism were represented among genes up-regulated in CTCs, with special reference to mitochondrial metabolism in CTC *vs.* LUNG class comparison, besides to general biosynthesis processes.

Genes down-regulated in CTC compared to solid lesions were enriched in GO terms mainly related to chromatin remodeling and regulation of transcription, consistently with the

considerable transcriptional reprogramming observed in previous analyses, and in terms related to tissue development/morphogenesis and regulation of cell adhesion and motility (Figure 4.2.10, Panel B).

Table 4.2.2. Gene Ontology (GO) term enrichment after pair wise class comparison using common differentially expressed (DE) genes in GEP1 and GEP2 experiments

classes	***GO terms for up-regulated genes***	***GO terms for down-regulated genes***
CTC *vs.* PT	Response to external stimuli/factors, cellular defense from pathogens, biosynthetic/metabolic processes, development/morphogenesis/differentiation	Chromatin/transcription regulation, development/morphogenesis/differentiation
CTC *vs.* LN	Development/morphogenesis/differentiation regulation of cell adhesion/migration, response to external stimuli/factors	Chromatin/transcription regulation, Development/morphogenesis/differentiation
CTC *vs.* LUNG	Mitochondrial metabolism, biosynthetic processes, development/differentiation, chromatin/transcription regulation	Chromatin/transcription regulation, development/differentiation
CTC *vs.* MDA-MB-231	Development/morphogenesis/differentiation metabolism, angiogenesis, biosynthetic processes, cell motility/migration	Development/morphogenesis/differentiation chromatin/transcription regulation, cell adhesion/motility
CTC *vs.* solid lesions	Development/morphogenesis/differentiation response to external stimuli/factors, cellular defense from pathogens, protein/ions localization, metabolism	Chromatin/transcription regulation, development/morphogenesis/differentiation cell adhesion/motility, cellular defense, metabolism
PT *vs.* LN	*No significantly DE genes*	*No significantly DE genes*
LUNG *vs.* PT	Development/morphogenesis/differentiation cell adhesion/motility	Metabolism, response to external stimuli/factors, development/morphogenesis/differentiation
PT *vs.* MDA-MB-231	Vasculature development, cell migration, regulation of secretion, protein modification	Cell migration/motility regulation, development/morphogenesis/differentiation
LN *vs.* LUNG	Glycolysis, biosynthetic/metabolic processes, development/morphogenesis/differentiation	Development/morphogenesis/differentiation
LN *vs.* MDA-MB-231	Cell adhesion/migration, angiogenesis, development/morphogenesis/differentiation ions transport, response to external stimuli/factors	Development/morphogenesis/differentiation migration, cell signaling
LUNG *vs.* MDA-MB-231	Development/morphogenesis/differentiation, immune system activation	Cell adhesion/migration, metabolism, development/morphogenesis/differentiation

Genes up-regulated in PT or LN *vs.* MDA-MB-231 were enriched in GO terms mainly related to vessel formation, crucial step of tumor progression, whereas those up-regulated in LUNG *vs.* MDA-MB-231 were again related to development and activation of defense mechanisms, processes required for adaptation to a new microenvironment (Table 4.2.2).

A list of genes of interest, some of them related to cell metabolism, epithelial-to-mesenchymal transition (EMT), transcriptional reprogramming and development, interaction with the tumor microenvironment or the acquisition of stemness/metastatic properties, including the expression trend in CTCs compared to solid lesions using merged data from GEP1 and GEP2, is reported in Table 4.2.3.

Table 4.2.3. Expression trend of genes of interest in CTC *vs.* solid lesions

gene	FC	Padj	biological process
HK2	-1.96	2.38E-08	metabolism
PDK1	-1.77	7.62E-06	
ACSS1	2.79	6.56E-07	
SNAI1	-1.00	0.97	EMT
SNAI2	-1.15	0.18	
VIM	1.05	0.23	
CDH1	-1.03	0.61	
CDH2	-1.06	0.16	
EPCAM	1.19	0.07	
PRRX1	-3.10	2.41E-08	
SPP1	1.03	0.81	microenvironment/metastasis
SPARC	-1.57	0.06	
PTPN11	-1.63	9.62E-07	stemness/metastasis
SOX2	-1.08	0.75	
YAP1	-1.60	2.34E-06	
HDAC10	3.13	2.30E-11	development
ELF3	4.18	6.66E-07	CTC cluster

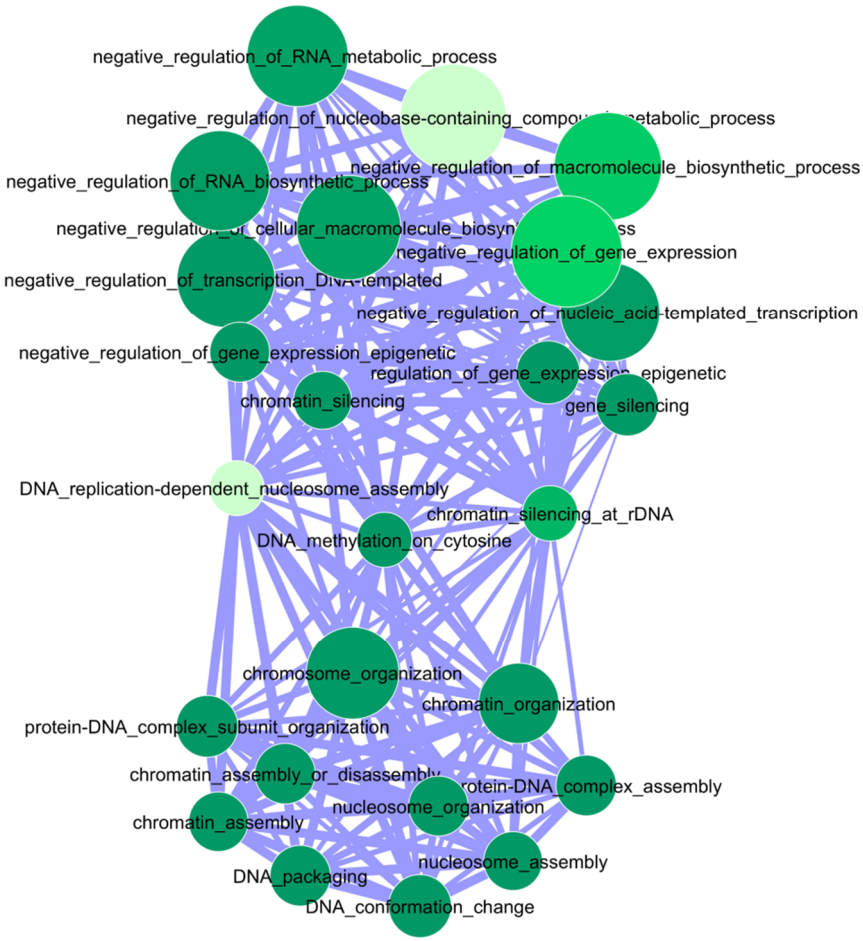
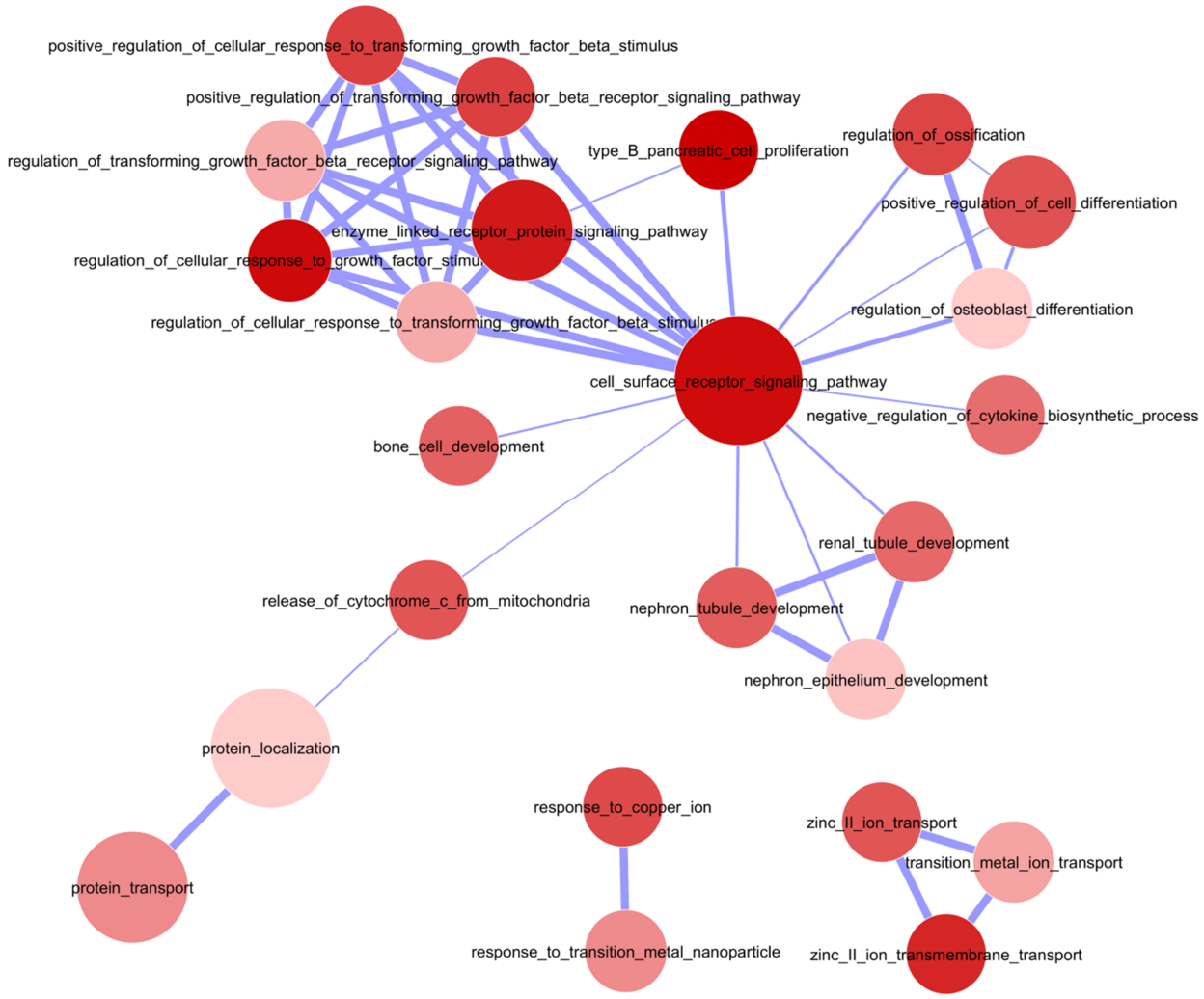


Figure 4.2.10. GO enrichment analysis of genes exclusively differentially expressed in CTCs from MDA-MB-231 CTC-model. Network representation of Gene Ontology (GO) terms enriched in the lists of genes found significantly down-regulated (green) or up-regulated (red, Figure 4.2.10 continued) in the comparison between CTC and solid lesions in both GEP1 and GEP2. Nodes represent significantly enriched GO terms. Their size is proportional to the number of genes annotated in the term and their color to their significance (darker color, smaller p-value). Nodes that share common genes are connected by an edge, with thickness proportional to the overlap coefficient (OC) between the two terms, calculated as $|A \cap B|/\min(|A|, |B|)$. Only terms with an OC ≥0.5 are shown.



(Figure 4.2.10. continued)

4.3. Role of TFF3 in dissemination and metastasis

4.3.1. Selection of candidate genes for functional validation studies with the MDA-MB-231 CTC-model

Metastasis is a complex biological process which successfully performs once tumor cells have acquired the ability to disseminate, survive in blood circulation, colonize distant organs and adapt to foreign microenvironments. GEP analyses of the CTC-model obtained from MDA-MB-231 cells orthotopic xenograft in immunocompromised mice revealed that CTCs have a unique transcriptional profile compared to PT, lymph-nodal metastases and lung metastases. In order to evaluate the biological contribute of genes differentially expressed in CTCs in the dissemination process, hypotheses were generated and criteria were defined to select candidate genes for functional validation *in vitro* and *in vivo* studies (the results of the analysis reported below refer to GEP1, the first experiment performed in 2013).

Focusing on the hematogenous dissemination phase, genes differentially expressed in CTC compared to the PT were firstly considered. Genes down-regulated ($log_2FC \leq -1.5$) or up-regulated ($log_2FC \geq 1.5$) in CTC compared to PT were 308 (106 with $P_{adj} < 0.0001$) and 291 (81 with $P_{adj} < 0.0001$), respectively. In order to preliminarily evaluate the involvement of candidate genes in biological processes related to metastasis using appropriate *in vitro* assays, and to provide the scientific rational that should justify the sacrifice of animals, a list of genes was selected also taking into account the relative expression level compared to the MDA-MB-231 parental cell line.

If considering the passage of the parental cell line 1) from an *in vitro* culture state 2) to different growth conditions after injection in mice, *i.e.* the PT take and growth, 3) to the hematogenous dissemination as three temporally distinct but biologically linked events, six lists of genes can be generated according to all possible categories coming out from multiple MDA231 vs. PT vs. CTC comparisons:

- “up/up”: if $log_2FC \geq 1.5$ in both paired class comparisons MDA-MB-231 *vs.* PT and PT *vs.* CTC
- “up/down”: if $log_2FC \geq 1.5$ in MDA-MB-231 *vs.* PT and $log_2FC \leq -1.5$ in PT *vs.* CTC
- “down/up”: if $log_2FC \leq -1.5$ in MDA-MB-231 *vs.* PT and $log_2FC \geq 1.5$ in PT *vs.* CTC
- “down/down”: if $log_2FC \leq -1.5$ in both paired class comparisons
- “notDE/up”: if $log_2FC \in]-1.5;1.5[$ in MDA-MB-231 *vs.* PT and $log_2FC \geq 1.5$ in PT *vs.* CTC
- “notDE/down" if log2FCϵ]-1.5;1.5[in MDA-MB-231 *vs.* PT and $log_2FC \leq -1.5$ in PT *vs.* CTC.

Genes significantly differentially expressed (P_{adj} <0.0001) between PT and CTC were then identified for all the aforementioned categories on the hypothesis that such genes could be involved in pathways that regulate the dissemination process. The results of this analysis are reported in Table 4.3.1.

Table 4.3.1. Number of genes differentially expressed according to multiple comparison categories among MDA-MB-231, PT and CTC classes

Categories of multiple MDA-MB-231 vs. PT and PT *vs.* CTC class comparisons	N° of genes (FC	N° of significantly (P_{adj}.<0.0001) DE genes in PT *vs.* CTC
up/up	6	**0**
up/down	51	**11**
down/up	50	**7**
down/down	11	**4**
notDE/up	252	**99**
notDE/down	229	**66**
TOTAL	*599*	***187***

For experimental reasons, *i.e.* to facilitate loss-of-function experiments, genes significantly differentially expressed between PT and CTC were analyzed according to the

relative expression in MDA-MB-231 *vs.* CTC class comparison categories. Therefore, after excluding genes not differentially expressed between MDA-MB-231 and PT (overall 165 if considering those expressed at comparable levels between CTC and both MDA-MB-231 or PT), subsequent analyses were focused on three biologically relevant categories, aforementioned as “up/up”, “down/up” and “down/down”.

Taken together, the results of these analyses (Table 4.3.2) revealed that:

- 16 genes were not significantly differentially expressed between MDA-MB-231 and CTC;
- 2 genes that underwent down-regulation in PT compared to MDA-MB-231 and were significantly up-regulated in CTC compared to PT were still significantly up-regulated in MDA-MB-231 compared to CTC (thus their expression levels fluctuated during tumor growth but remained substantially lower in CTC compared to the parental cell line),
- 4 genes up-regulated during tumor growth compared to MDA-MB-231 and up-regulated during dissemination compared to PT were still significantly up-regulated in CTC compared to MDA-MB-231 (thus their expression levels were substantially higher in CTC compared to the parental cell line).

On the hypothesis that the PT represents a biological filter where, during tumor growth, few cells committed to dissemination undergo positive selection and that genes up-regulated in CTC *vs.* PT could be involved in intravasation and survival in blood, the attention was focused on genes expressed in MDA-MB-231 which underwent significant down-modulation after injection in mice and were re-activated in disseminated cells (listed under the “up/down” category and marked in grey in Table 4.3.2).

Before performing gene modulation experiments, the expression level of *TFF1*, *TFF2*, *TFF3*, *ELF3* and *NR4A1* was evaluated by real-time PCR in MDA-MB-231 cells. *TFF2*, *TFF3* and *ELF3* were expressed at high level ($C_t \in [29.9\text{-}24.0]$), whereas *TFF1* and *NR4A1*, were expressed at intermediate level ($C_t \in [32.9\text{-}30.0]$) (Figure 4.3.1, Panel A).

Table 4.3.2. List of genes obtained after multiple comparison analysis (MDA-MB-231 *vs.* PT and PT *vs.* CTC) and classified according to the expression level between CTC and MDA-MB-231

up/down	down/up	down/down	notDE/up	notDE/down
TFF2	*LST-3TM12*	*NCF1*	*C13ORF23*	*NR4A3*
TFF1	*SNCAIP*	*CRIP2*	*AOX1*	*LY6D*
NR4A1	*NAV3*	*C1QB*	*RN7SK*	*MIR196A2*
ELF3	*SNORD55*	*UBE2NL*	*SNORA45*	*PTGS1*
FLJ45445	*SNORA51*		*INO80D*	*CRYGB*
MMP28	*SNORA47*		*KIAA1632*	*BCAS1*
TFF3	*SNORD95*		*AGPAT5*	*CDH8*
ACSS1			*ZNF407*	*OR1L4*
INHBB			*JARID2*	*H19*
SPINK4			*SNORA53*	*SYN1*
IRX5			*KIAA2018*	*CITED1*
			SNORA25	*CDH29*
			SACS	*C6ORF27*
			CTNNB1	*RRAD*
			CRYBG3	*ADAMTS14*
			C2ORF86	*CD14*
			ZNF460	*MAMDC2*
			…	…

genes not differentially expressed (DE) in MDA-MB-231 *vs.* CTC

genes up-regulated in MDA-MB-231 *vs.* CTC

genes down-regulated in MDA-MB-231 *vs.* CTC

genes up- or down-regulated in CTC compared to both PT and MDA-MB-231

Considering that forced expression of *TFF1* was already proven to increase cell migration and invasion in MDA-MB-231 cells and other BC cell lines (Buache et al., 2011), and that *NR4A1* was expressed at low level and was proven to be involved in metastasis in MDA-MB-231 xenograft models (Zhou et al., 2014), gene silencing experiments were performed to down-modulate the expression of *TFF2*, *TFF3* and *ELF3*.

As the silencing of *ELF3* and *TFF2* genes failed when using different commercially available siRNA molecules (data not shown), functional validation experiments were performed to evaluate the role of *TFF3* in proliferation, migration, invasion, vascular mimicry,

dissemination and metastasis. The distribution of signal intensities of *TFF3* in GEP1 and GEP2 experiments are reported in Figure 4.3.1 Panel B. The overall *TFF3* expression trend was confirmed by moderated t-statistics (ANOVA P_{adj} <0.0001).

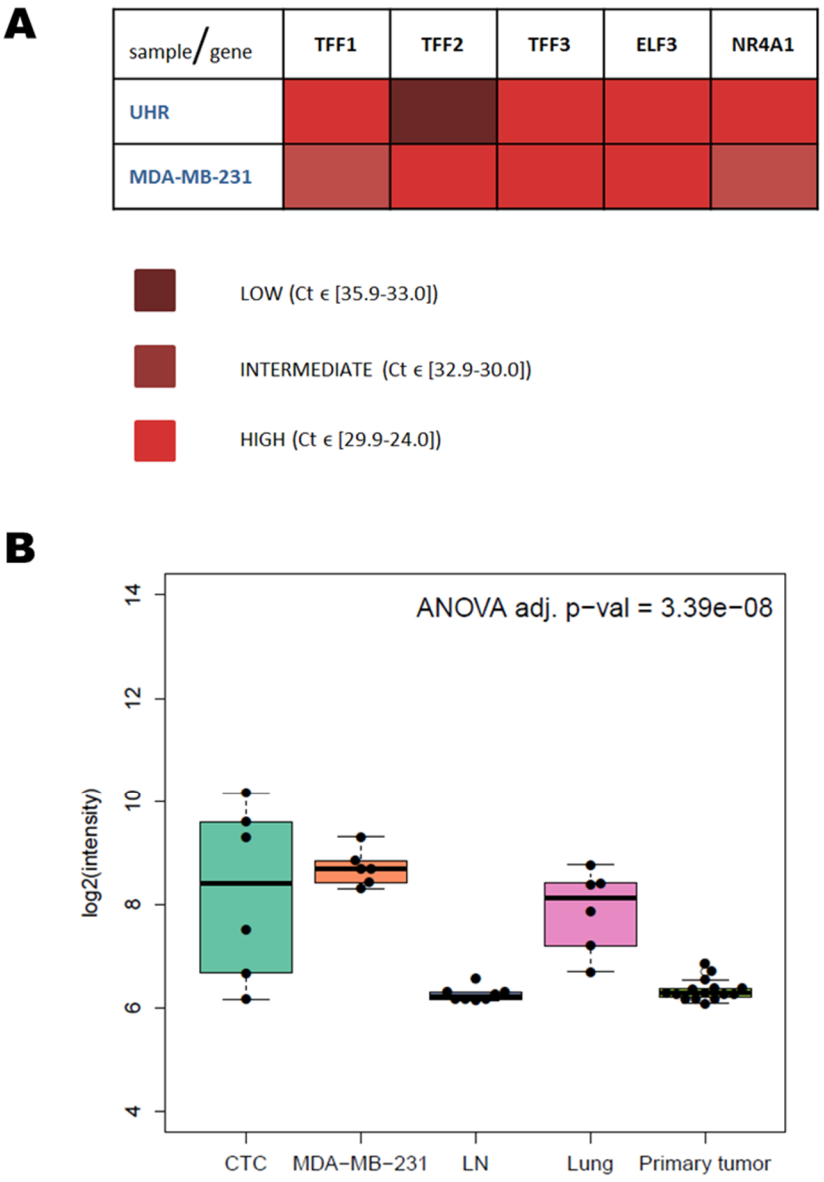


Figure 4.3.1. Expression of candidate genes in MDA-MB-231. Color code (**A**) represents the expression levels of *TFF1*, *TFF2*, *TFF3*, *ELF3* and *NR4A1* genes in MDA-MB-231 cells, ranked according to C_t values obtained by qPCR analysis. Box plot (**B**) represents the expression levels of *TFF3* in CTC, MDA-MB-231, LUNG, LN and PT samples using data from GEP1 and GEP2.

4.3.2. *TFF3* transient gene silencing impairs MDA-MB-231 migration but not proliferation

Polymer-mediated siRNA delivery and *TFF3* silencing efficiency were firstly assessed in order to estimate the number of cells on the whole population able to take the siRNA molecule and the protein knock-down efficiency during time. The transfection efficiency after 48 hours was ~75% (Figure 4.3.2, Panel A) and a stepwise decrease in *TFF3* gene expression and in TFF3 peptide secretion in the conditioned medium was observed in a time-course experiment (Figure 4.3.2, Panels B and C). The best knock-down efficiency was reached 6 days after transfection (12% relative optical density, O.D., *vs.* cells transfected with a negative control siRNA).

Preliminary *in vitro* assays were performed in order to understand if TFF3, identified as a CTC-related gene in the MDA-MB-231 CTC-model, is involved in cell proliferation or if it exerts specific role in biological processes peculiar to the metastatic process. The involvement of TFF3 in MDA-MB-231 proliferation and migration was assessed by MTS and Boyden chamber migration assays, respectively. TFF3 transient knock-down did not affect cell proliferation in a time-course experiment (Figure 4.3.2, Panel D), but significantly impaired cell migration ability as an 84%±7% mean±SD reduction was obtained in silenced *vs.* control cells ($P = 0.0031$, two-tailed Student's *t* test; Figure 4.3.2, Panel E).

These preliminary results provided the scientific rationale to assess the metastatic potential of *TFF3* knock-out MDA-MB-231 cells in mouse models.

4.3.3. MDA-MB-231 cell heterogeneity influences the outcome of *in vivo* validation studies after CRISPR/Cas9-mediated *TFF3* knock-out

MDA-MB-231 cells were transfected with vectors containing a guide RNA for *TFF3* gene deletion mediated by the CRISPR/Cas9 system. Cells were seeded in order to obtain single clones, and after two weeks in culture 1,018 out of 4,233 screened wells contained proliferating cells. Despite the low plasmid transfection efficiency (about 1.3%, Figure 4.3.3,

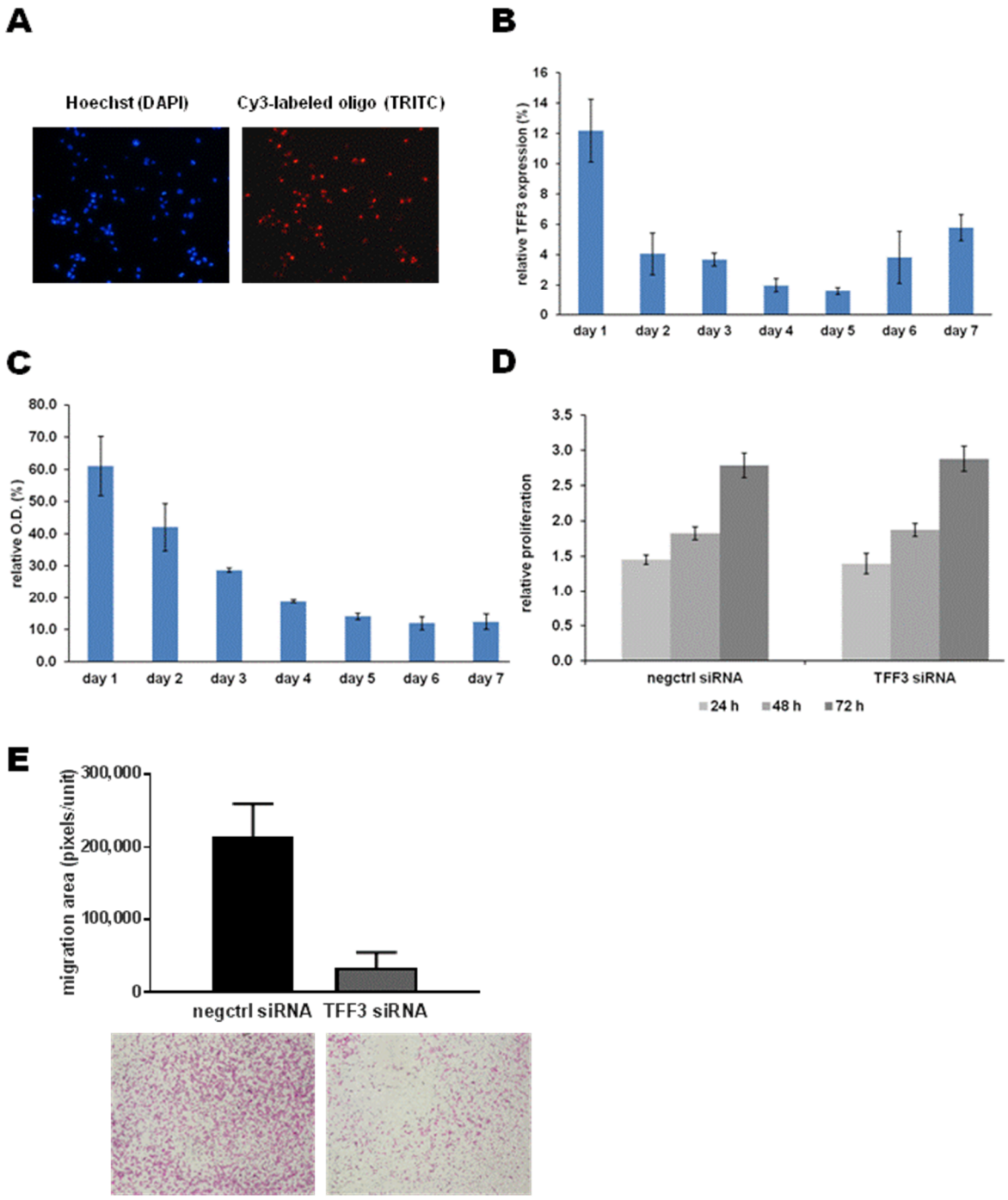


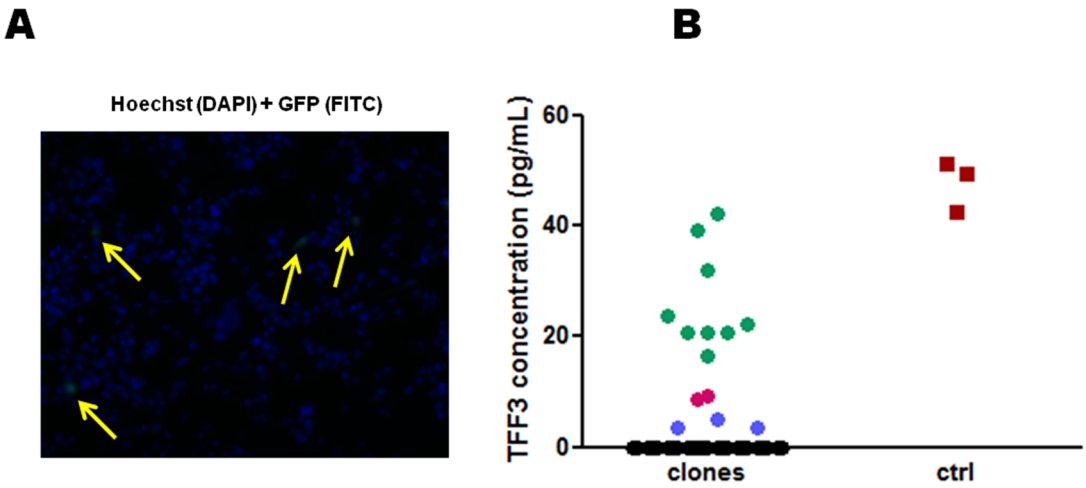
Figure 4.3.2. siRNA-mediated *TFF3* silencing effect on MDA-MB-231 cell proliferation rate and migration ability. Pictures (**A**) acquired by DAPI (blue) and TRITC (red) filters (200X total magnification) represent MDA-MB-231 cells transfected for 48 hours with a Cy3-labeled oligo; nuclei were stained with Hoechst 33342. Relative TFF3 levels measured by (**B**) qPCR and (**C**) ELISA assay in MDA-MB-231 cells transfected with a *TFF3* siRNA *vs.* cells transfected with a negative control siRNA. Bar charts represent the mean±SD relative gene expression or relative optical density (O.D.) obtained from three biological replicates, at different time points. Proliferation rate (**D**) measured by MTS assay. Bar chart represents the mean±SD relative O.D. (normalized to O.D. measured 12 hours after seeding) of three independent experiments performed in triplicate, measured at different time points. Migration ability (**E**) assessed by Boyden chamber assay. Bar chart represents the mean±SD migration area (pixels per unit) from three independent experiments performed in sestuplicate. Images below the graph represent migration areas obtained after 36 hours (40X total magnification).

Panel A), the CRISPR/Cas9 system allowed obtaining complete TFF3 deletion in 84% of cases as TFF3 concentrations in conditioned media were a) undetectable in 76/90 clones, b) 88.54%±19.22% mean±SD relative expression *vs.* control parental cells in 9/90 clones, c) on average 34% *vs.* control parental cells in 2/90 clones, d) under the mean minimum detectable dose indicated by the manufacturer of the ELISA assay (6.43 pg/mL) in 3/90 clones (Figure 4.3.3, Panel B). Validation of gene knock-out by Sanger sequencing was performed in 5 wild-type (transfected with control vector) and 8 knock-out clones; the results of the nucleotide sequence alignment of the three *TFF3*^{KO} clones injected in animals for *in vivo* functional experiments were reported in Figure 4.3.3, Panels C-E, for instance.

The proliferation rate of 13 clones was assessed by MTS assay. High variability among clones emerged as a stepwise increase in relative proliferation was observed in two wild-type (#02 and #04) and 6 knock-out clones (#25, #27, #29, #70, #76 and #91), similarly to the parental cell, whereas a steady state was observed after two days in the 3 out of 5 (#01, #03 and #05) and 2 out of 8 (#22 and #27) remaining wild-type and knock-out clones, respectively (Figure 4.3.4, Panel A).

In parallel, migration assays were performed in order to verify the biological behavior observed in previous gene silencing experiments and select cell clones that better reproduce the behavior of the parental cells. Heterogeneity among clones was observed also in these experiments, as the calculated migration area ranged from roughly 800 to 460,000 pixels per unit, irrespective of the experimental group (Figure 4.3.4, Panel B). Consistently with the slower proliferation rate, wild-type clones derived from single MDA-MB-231 cells transfected with control vector showed significantly weaker migrating potential compared to the parental cells, as migrating areas ranged from 7% to 73% of control (p-values 0.0005, 0.2179, 0.0409, 0.0001 and <0.0001, two-tailed Student's *t* test, respectively for #01, #02, #03, #04 and #05). Unexpectedly, knock-out clone #72 showed a 2-fold higher migrating potential compared to the parental cells ($P = 0.0188$, two-tailed Student's *t* test), and clone #91 did not differ from the wild-type clone #01 (70,000 pixels per unit). However, all of the other

knock-out clones confirmed the impairment in cell motility observed in gene silencing experiments.



C

```
Query  160  TTGTTTGCACAGCTGCTCTGGATTGTTTGCTTGGGGAAGG-TCTCCCTCCTGGCTGTGG-  217
             ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Sbjct  153  TTGTTTGCACAGCTGCTCTGGATTGTTTGCTTGGGGAAGGCTCTCCCTCCTGGCTGTGGC  94
```

D

```
Query  160  TGCAGAGCGC-------------------GTGGGCTCCGGGACGCAGCTCAGGACTCGCT  200
             ||||||||||                   ||||||||||||||||||||||||||||||
Sbjct  244  TGCAGAGCGCTCTGGCAGCCATGACCACCGTGGGCTCCGGGACGCAGCTCAGGACTCGCT  185
```

E

```
Query  579  AGGCAGTTTTTCTCTCCCAAAGGGGAGAGGTCA  611
             || ||||||||||||||||||||||||||||||
Sbjct  55   AGACAGTTTTTCTCTCCCAAAGGGGAGAGGTCA  23
```

```
Query  161  TGCAGAGCGGCTCTGGCAGCCATGACCACCGTGGGCTCCGGGACGCAGCTCAGGACTCGC  220
             ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  244  TGCAGAGCG-CTCTGGCAGCCATGACCACCGTGGGCTCCGGGACGCAGCTCAGGACTCGC  186
```

Figure 4.3.3. Results of *TFF3* knock-out experiments with the CRISPR/Cas9 system. Composite picture (**A**) acquired by DAPI (blue) and FITC (green) filters (200X total magnification) represents MDA-MB-231 cells transfected with a pSpCas9(BB)-2A-GFP control plasmid for 48 hours. Nuclei were stained with Hoechst 33342; yellow arrows indicate GFP-expressing cells. Dot plot (**B**) represents the concentration of TFF3 assessed by ELISA test: green, pink, blue and black dots represent cell clones with high, medium, under the minimum detectable dose and undetectable TFF3 expression compared to non transfected control cells ("ctrl"). Output of BLAST program (**C-E**) represents the results of nucleotide alignment with the reference gene sequence for clones (**C**) #22, (**D**) #25 and (**E**) #70 (sequencing of 2 different PCR bands).

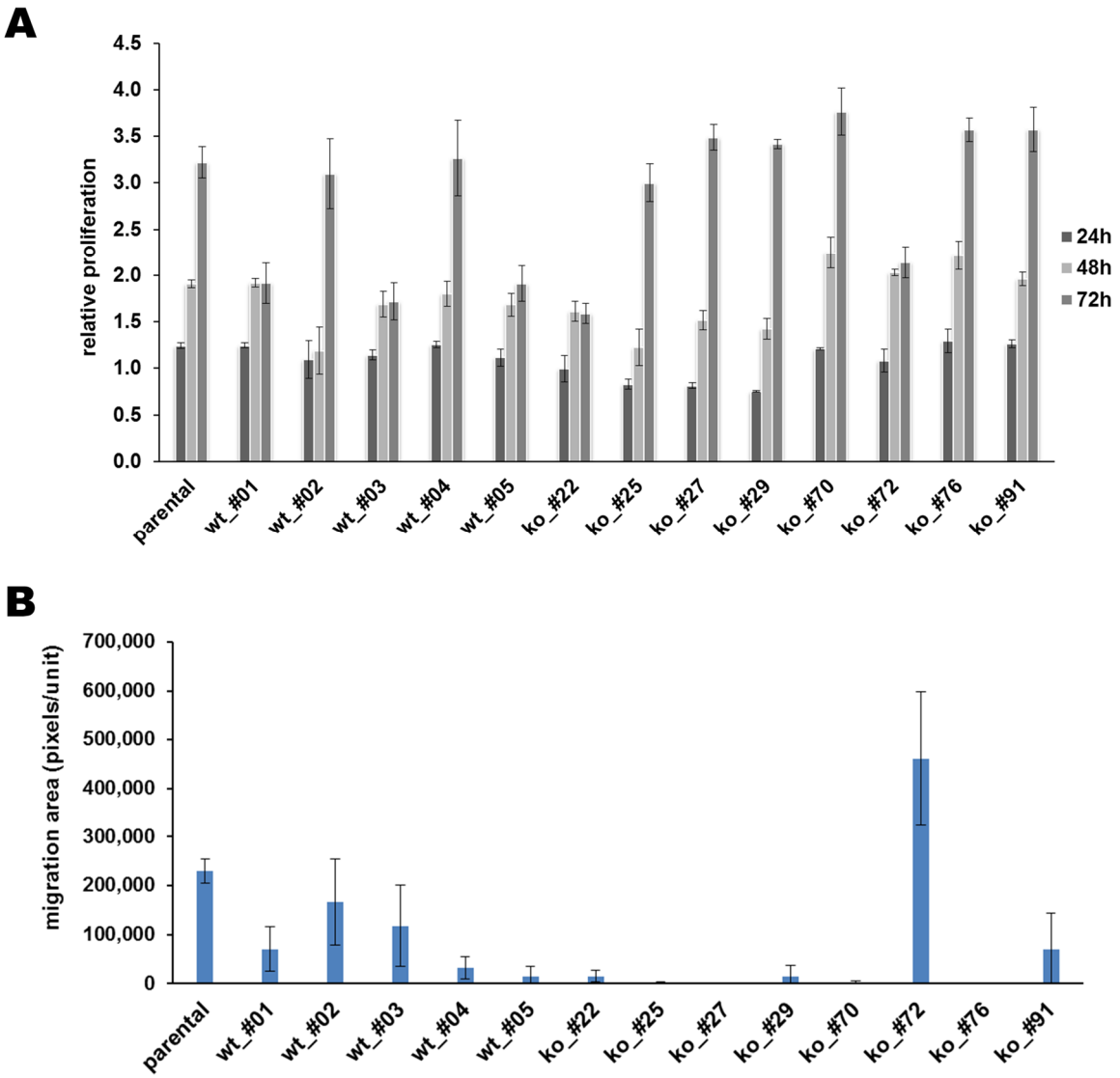


Figure 4.3.4. Proliferation and migration of MDA-MB-231 cell clones after CRISPR/Cas9-mediated *TFF3* knock-out. Proliferation rate (**A**) of untreated MDA-MB-231 cells, five *TFF3*^{WT} and eight *TFF3*^{KO} clones, measured by MTS assay. Bar chart represents the mean±SD relative O.D. (normalized to O.D. measured 12 hours after seeding) of three experimental replicates, measured at different time points. Migration ability (**B**) assessed by Boyden chamber assay 36 hours after seeding. Bar chart represents the mean±SD migration area (pixels per unit) of six experimental replicates.

Considering that MDA-MB-231 cells are genetically unstable and that the prolonged clone selection protocol and the amplification steps entailed 7 to 9 *in vitro* passages, all 13 clones underwent authentication test. Wild-type #04 and #05 clones did not pass quality control by STR profiling as they did not match for 1 out of 8 analyzed loci. Therefore, they

were excluded from the following experiments. All knock-out clones passed the authentication test. Among knock-out clones, #27, #29 and #76 were excluded as they displayed a proliferating rate less similar to those of wild-type clones, in addition to #72 and #91, which displayed unusual increased migrating potential. Therefore, #22, #25 and #70 were chosen for *in vivo* dissemination and metastasis formation experiments, in comparison with the group of selected wild-type clones (#01, #02 and #03). In the matter of question, migrating ability was still significantly lower compared to wild-type clones (P <0.05, two-tailed Student's t test), except when comparing wild-type #03 with knock-out #22 and #25 clones (P = 0.06, two-tailed Student's t test).

Cells were injected according to the same experimental conditions used for MDA-MB-231. Differently from the parental cells and despite results obtained by MTS proliferation assay, all clones displayed longer latency time before tumor take and a lower growth rate after injection in animals. Animals injected with *TFF3*^{WT} and *TFF3*^{KO} clones which showed comparable growth rate (*i.e.*, #03 and #22, respectively, Figure 4.3.4, Panel A) were sacrificed 100 days after tumor cells implant, when they had reached about 1 g of total tumor mass. Growth rate of wild-type and knock-out cells in mice was not significantly different. CTCs and CTM were quantified by direct analysis on cytological samples and the presence of metastases at lung and LNs was assessed by IHC analysis on 20 consecutive sections per organ. CTCs were not detectable in the blood of the three animals injected with the knock-out clone #22, whereas 9 CTC/mL in 1 out of 3 samples were observed in the blood sample of mice injected with the wild-type clone #03 (Figure 4.3.5, Panel B). As expected, in consideration of the results obtained in *in vitro* assays, discordant data on metastasis formation were obtained as wild-type clones failed to form metastases in a similar extent compared to mice injected with parental cells, and developed instead just single scattered metastatic cells or small cell clusters (5-10 cells per cluster). On the contrary, knock-out clones, although the extent of metastases was lower than those observed in mice injected with the parental cells, showed larger tumor clusters (from 2 to 200 cells per cluster) compared to the wild-type clone in all cases.

A similar behavior was observed also in LN FFPE sections, which were metastasis-positive in knock-out but not in wild-type cases.

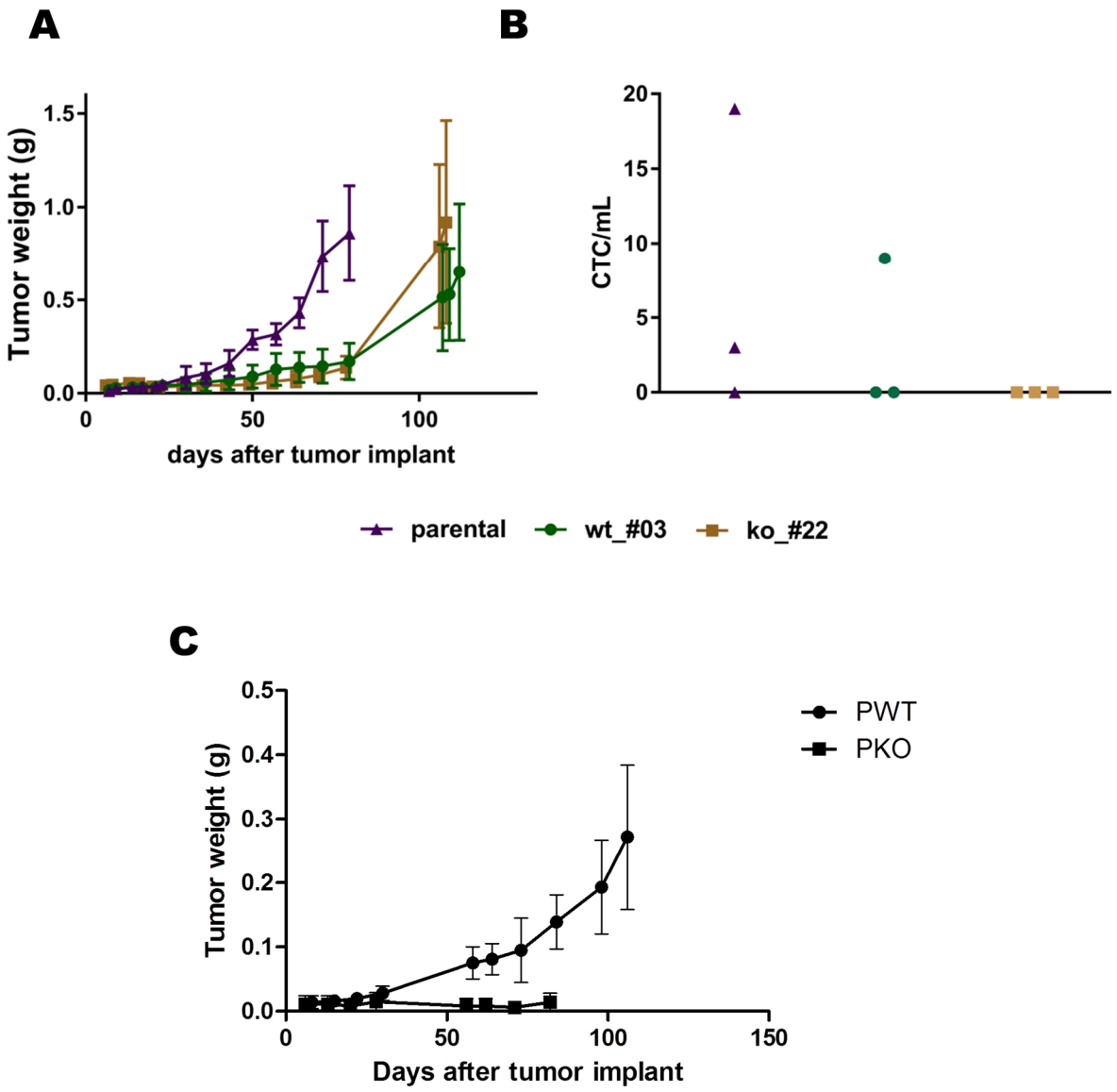


Figure 4.3.5. Tumor growth rate and CTC load of MDA-MB-231 *TFF3*^{WT} and *TFF3*^{KO} clones. Line graph represents the mean±SD of the sum of the masses of two nodules from three NOD scid mice injected with (**A**) MDA-MB-231 parental cells, the *TFF3*^{WT} #03 and the *TFF3*^{KO} #22 MDA-MB-231-derived clones, and dot plot (**B**) represents the distribution of CTCs isolated at animal sacrifice. Line graph (**C**) represents the mean±SD of the sum of the masses of two nodules from three NOD scid mice injected with a pool of the three *TFF3* wild-type (#01, #02 and #03; "PWT") and the three knock-out (#22, #25 and #70; "PKO") clones.

In an attempt to restore the complexity of the MDA-MB-231 cell population, mice were injected with a pool of the three wild-type or knock-out clones. In this case, despite a faster growth rate in the wild-type compared to the knock-out experimental group, both pools of cells were not able to form tumor nodules with a mass higher than 0.5 g after 80 days from injection (Figure 4.3.5, Panel C).

Selection of single cell clones from the MDA-MB-231 cells did not allow drawing conclusions on the biological role of TFF3 due to the high heterogeneity and to the discordant results in terms of CTC release and pulmonary metastases formation.

4.3.4. TFF3 is involved in MDA-MB-231 migration and invasion, but not proliferation and vascular mimicry: confirmation experiments after stable knock-down

MDA-MB-231 cell heterogeneity influenced the outcome of *in vivo* validation studies after CRISPR/Cas9-mediated *TFF3* knock-out. In order to overcome this problem, an experiment was performed to obtain *TFF3* stable knock-down cells by lentivirus-mediated delivering. At first, infection efficiency was assessed and estimated to be ~95% (Figure 4.3.6, Panel A). Two out of the three shRNA vectors caused ~80% stable knock-down of the secreted peptide compared to cells infected with lentivirus delivering a control shRNA vector (Figure 4.3.6, Panel B).

Proliferation and migration assays were repeated for confirmation. The proliferation rate was comparable in both control and TFF3 knock-down cells, although slight but not significantly lower than MDA-MB-231 parental cells (Figure 4.3.5, Panel C). Migrating ability was prevalently impaired after stable knock-down (99% and 90% reduction, $P = 0.0068$ and $P = 0.0109$, two-tailed Student's *t* test, in sh_#1 and sh_#2, respectively, compared to control cells; Figure 4.3.6, Panel D), thus confirming the results obtained with the two different gene silencing and knock-out systems.

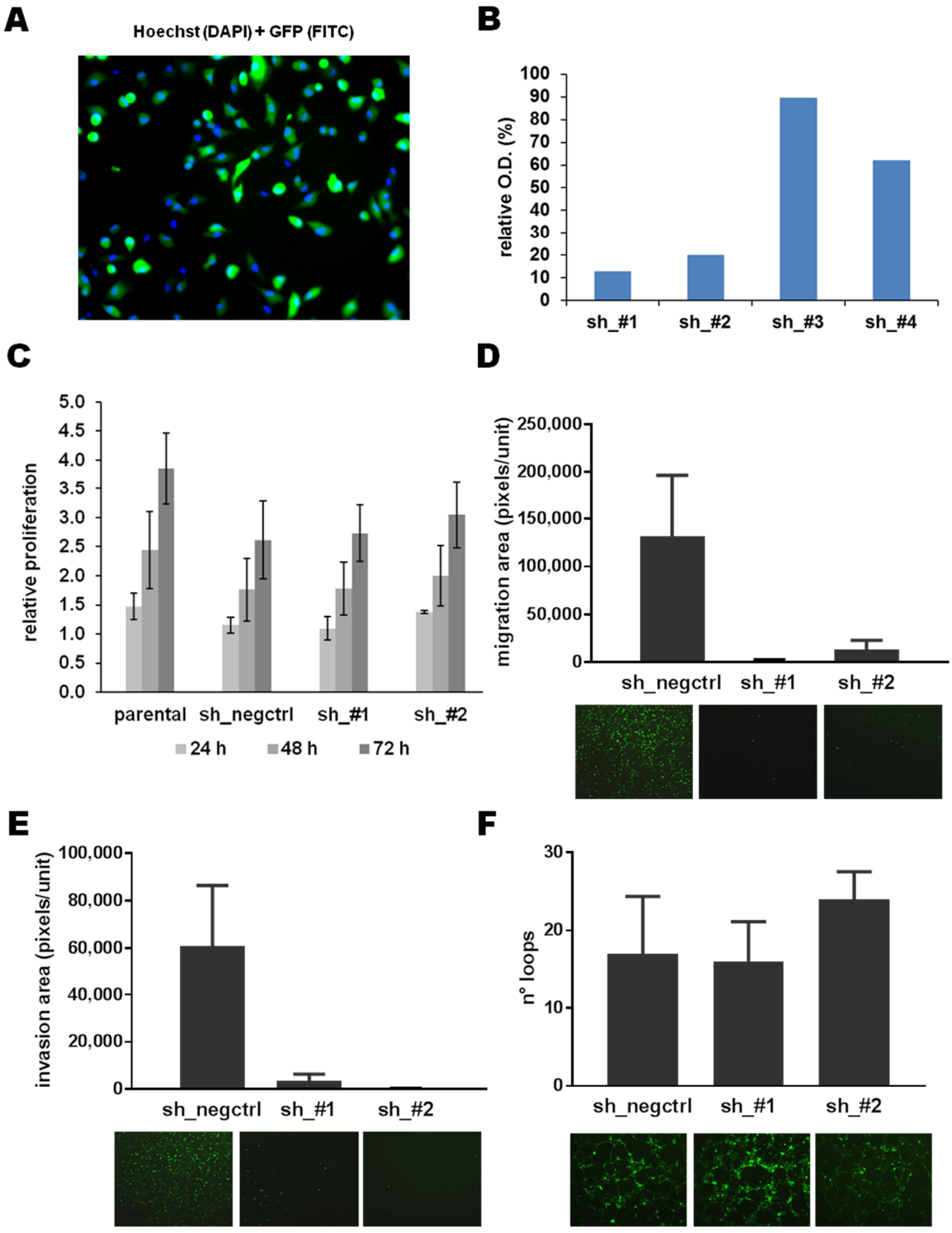


Figure 4.3.6. Proliferation, migration, invasion and vascular mimicry of MDA-MB-231 cells after TFF3 stable knock-down. Composite picture (**A**) acquired by DAPI (blue) and FITC (green) filters (200X total magnification) represents MDA-MB-231 cells infected for 48 hours with lentiviral particles delivering pGFP-C-shLenti control vector; nuclei were stained with Hoechst 33342. Bar chart (**B**) represents the relative expression (optical density) of secreted TFF3 measured in media conditioned by MDA-MB-231 cells infected with lentiviral particles delivering 4 different shTFF3 *vs.* cells transfected with a control vector. Proliferation rate (**C**) of untreated MDA-MB-231 cells, and cells infected with lentiviral particles delivering two sh*TFF3* ("sh_#1" and "sh_#2") or negative control ("sh_negctrl") vectors, measured by MTS assay. Bar chart represents the mean±SD relative O.D. (normalized to O.D. measured 12 hours after seeding) of three independent experiments performed in

(Figure 4.3.6. continued) triplicate, measured at different time points. Migration (**D**) and invasion (**E**) ability assessed by Boyden chamber assay. Bar chart represents the mean±SD migration or Matrigel invasion area (pixels per unit) of four and five experimental replicates, respectively. Images below the chart (40X total magnification, acquired by FITC filter) represent cells migrating across the nude or Matrigel-coated Transwell barrier 24 hours after seeding. Vascular-mimicry ability (**F**) assessed by Matrigel-based tube formation assay. Bar chart represents the mean±SD number of loops from three experimental replicates. Images below the chart (40X total magnification, acquired by FITC filter) represent cells forming vessel-like networks.

Invasion assay was also performed to corroborate the role of TFF3 in key steps of the metastatic cascade. Beyond migration, TFF3 was proven to be involved also in Matrigel invasion ability as the area invaded by TFF3 knock-down cells was significantly smaller compared to control (6% and 0.3% of control, $P = 0.0014$ and $P = 0.0010$, two-tailed Student's *t* test, respectively in sh_#1 and sh_#2 cells; Figure 4.3.6, Panel E).

Finally, TFF3 knock-down failed to impair the ability of tumor cells of non-vascular origin like MDA-MB-231 to acquire vascular cell characteristics and form cell-lined channels (*i.e.*, vascular mimicry), as the number of loops in knock-down was comparable to those in control cells (Figure 4.3.6, Panel F).

This experimental setting will provide the advantage of performing validation studies on a cell population highly representative of the whole parental cell line. To this aim, *in vivo* experiments are running using TFF3 knock-down cells in order to draw conclusions on the role of TFF3 in dissemination and metastasis formation in the MDA-MB-231 CTC model.

4.4. CTC longitudinal analysis in breast cancer case series: a comparison of two methods

4.4.1. Technical and clinical implications of CTC monitoring by AdnaTest

Longitudinal analysis of CTCs in specific clinical settings may provide information on response to therapy in real-time and may allow identification of biomarkers, either intended as CTC status or CTC count or CTC-specific molecular markers, to stratify patients at high risk of relapse or progression. Since different kinds of messages can be obtained according to the method used to interrogate CTCs, in the present study CTC monitoring was carried out using either a positive-selection based method, the AdnaTest EMT-1/Stem and EMT-2/Stem CellSelect and Detect kits, and a size-based one, the ScreenCell® Cyto kit, followed by multiplex RT-PCR and cytological analysis, respectively. The study was performed on independent blood samples from two series of patients under neoadjuvant therapy and two series of patients under treatment for metastatic disease. Part of the results presented here was published in Fina et al., 2015b.

Preliminary spiking experiments were performed using the luminal MCF7 and basal MDA-MB-231 BC cell lines, which express EpCAM and MUC1 on the cell surface at high and negligible level, respectively, in order to test the technical and biological sensitivity of AdnaTest. The AdnaTest EMT-1/Stem CellSelect kit (hereafter referred to as “EMT1”), which consists of immunomagnetic beads for isolation of EpCAM- and MUC1-positive cells, detected at least 5 MCF7 cells spiked in 5 mL of blood from a healthy donor, but failed to detect MDA-MB-231 cells (Figure 4.4, Panel A). After implementation of the panel of surface markers for CTC enrichment with additional antibodies against ERBB2 and EGFR (information kindly provided by the manufacturer of AdnaTest EMT-2/Stem CellSelect kit, hereafter referred to as “EMT2”), the approach proved to be adequate for the isolation of both epithelial- and mesenchymal-like cells, as it allowed also the isolation of at least 5

MDA-MB-231 cells, detectable by RT-PCR for the mesenchymal marker cMET, in a further EMT2 *vs.* EMT1 comparison test (Figure 4.4, Panel B).

Two 5 mL-blood samples withdrawn from a series of 5 female healthy donors were analyzed using in parallel both AdnaTest kits, in order to verify the specificity of the method and to define a cut-off value for *ACTB*, in this case ≥0.70 ng/μL, as quality control of the entire methodological approach (Table 4.4.1).

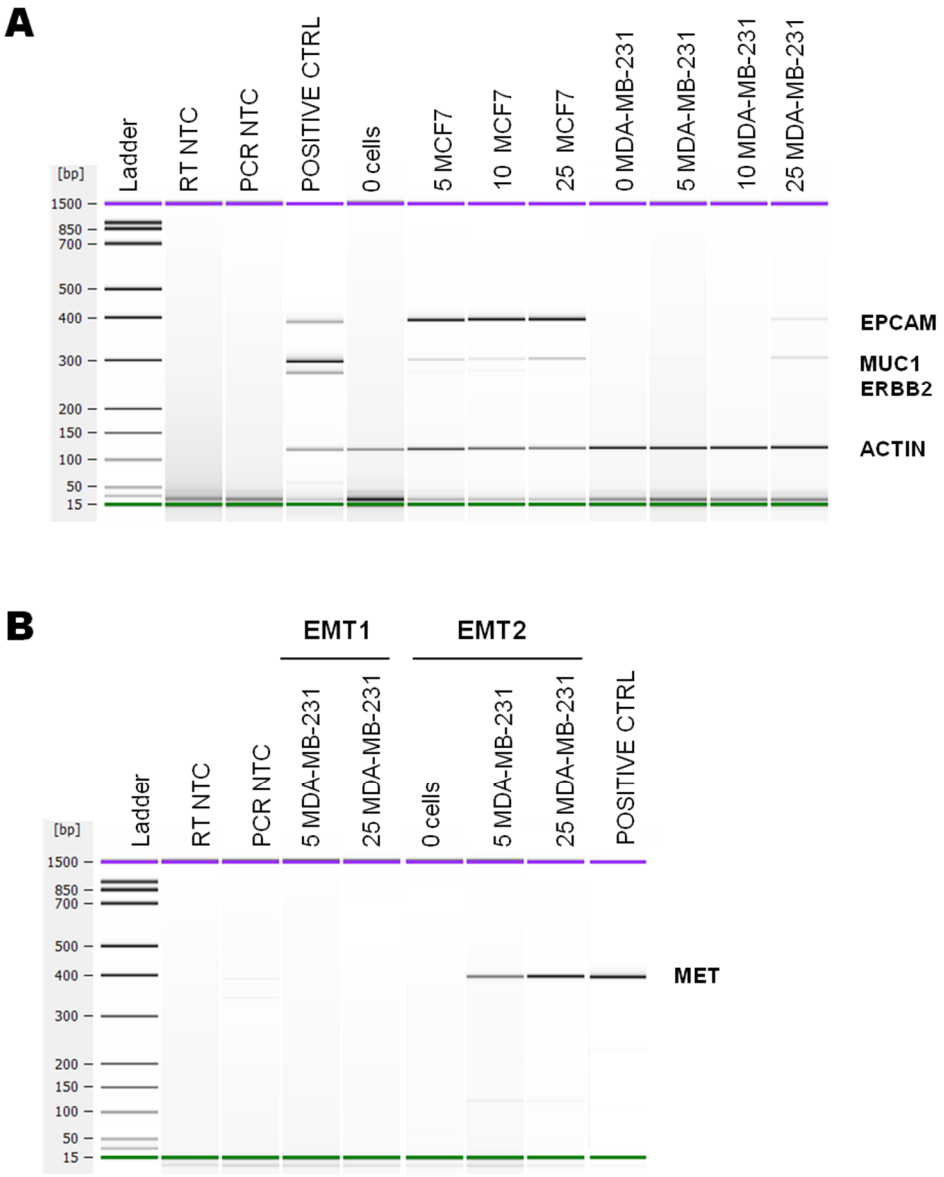


Figure 4.4.1. Biological sensitivity of AdnaTest EMT1 and EMT2 kits. Gel-like images of capillary electrophoresis represent optical view of signal intensities of (**A**) *EPCAM*, *MUC1* and *ERBB2* PCR products from MCF7 and MDA-MB-231 cells spiked in the blood of healthy donors and captured by the EMT1 kits, and (**B**) *MET* PCR product from MDA-MB-231 cell spiked in the blood of healthy donors and captured by the EMT1 or the EMT2 kit.

CTC monitoring (started in 2012) in two BC cases series including patients undergoing neoadjuvant therapy or treatment for metastatic disease, was performed with EMT1 and in parallel, once commercially available (in 2014), with EMT2 kits. CTC positivity was assessed in 21 blood samples (3 before starting neoadjuvant treatment, 8 during treatment, 8 before surgery and 2 after surgery) withdrawn from 13 BC patients clinically diagnosed as non-metastatic (M0), and in 9 samples (2 before starting first-line systemic treatment, 4 during therapy, 2 at progression and 1 before starting second-line systemic therapy) withdrawn from 6 women with metastatic disease (M+).

Table 4.4.1. Analysis of blood samples by AdnaTest in healthy donors

EMT kit and donor ID	*Breast PCR*				*EMT PCR*				*Stem PCR*
	EPCAM	MUC1	ERBB2	ACTB	PIK3CA	AKT2	TWIST1	ACTB	ALDH1
EMT1 d3	0.00	0.00	0.00	1.82	0.00	0.04	0.00	0.90	0.00
EMT2 d3	0.00	0.00	0.00	2.98	0.03	0.03	0.00	1.58	0.00
EMT1 d4	0.00	0.00	0.03	3.48	0.06	0.08	0.01	1.58	0.00
EMT2 d4	0.05	0.00	0.00	3.34	0.09	0.08	0.00	1.61	0.00
EMT1 d5	0.00	0.00	0.00	1.54	0.01	0.03	0.00	0.71	0.00
EMT2 d5	0.00	0.00	0.02	2.58	0.00	0.06	0.00	0.81	0.00
EMT1 d6	0.00	0.00	0.00	3.23	0.13	0.07	0.00	1.33	0.00
EMT2 d6	0.00	0.00	0.00	3.93	0.18	0.07	0.00	1.45	0.00
EMT1 d7	0.00	0.00	0.00	1.95	0.00	0.02	0.01	0.88	0.00
EMT2 d7	0.00	0.00	0.00	1.77	0.00	0.00	0.00	0.77	0.00

Marker concentrations are expressed as ng/µL

As expected on the basis of spiking experiments using cell lines with marked epithelial or mesenchymal-like phenotype, EMT2 was able to detect CTCs in a higher number of samples compared to EMT1, 14/30 (47%) *vs.* 4/30 (13%), respectively, when considering the overall case series, thus providing further evidence for the existence of different CTC subpopulations in the same patient. Consistently, the overall agreement between EMT1 and

EMT2 on CTC status was poor (Cohen's κ = 0.159; 95%CI (-0.159-0.417)). Only 60% of samples were concordantly rated for CTC status and the discordant samples were enriched in EMT1-/EMT2+ cases (11 EMT1-/EMT2+ versus 1 EMT1+/EMT2-). Similar results were observed also after stratification by clinical setting: a 4-fold and a 3-fold increase in CTC detection rates were obtained in M0 and M+ cases, respectively, when using EMT2 compared to EMT1 kit. Therefore, improvement in CTC detection mainly impacts the outcome of studies with patients at early stage of disease and to a lesser extent those involving metastatic patients.

Since the improvement in CTC positivity rate does not solely depend on the efficacy of the enrichment method, but also relies on the specificity and representativeness of the markers used for CTC detection at transcript level, the effect on CTC positivity when using different panels of genes was also assessed. CTC positivity was defined including genes related to the activation of the epithelial-to-mesenchymal transition process (EMT), as *TWIST1*, *PIK3CA and AKT2*, and the stemness-related gene *ALDH1A*, beside to the classical epithelial- and breast tumor-associated markers *EPCAM*, *MUC1* and *ERBB2* (Table 4.4.2). Overall, independently from the kit used for CTC enrichment, CTC detection rates increased when including the panel of EMT- and stemness-related genes. Interestingly, the expression of epithelial/breast-tumor markers was sufficient to detect CTCs in the majority of patients with advanced stage of disease, since only slight changes could be observed when the other two panels of genes were included. On the contrary, in M0 cases the implementation of the EMT-gene panel determined a switch from a negative to a positive CTC status (when CTC enrichment was performed with EMT1) and a 6-fold increase in positivity rates (when the enrichment was performed with EMT2), compared to positivity rates assigned according to the expression of the classical epithelial/breast tumor markers only. Higher increase in CTC positivity was still observed in M0 cases when including also *ALDH1A*, thus obtaining a 2-fold and 1.3-fold increase for EMT1 and EMT2 capture kits, respectively, with an overall 8-fold increase compared to CTC status assessed by using epithelial/breast-tumor genes only (EMT2 kit).

On the basis of these results, heterogeneity in CTC population was hypothesized when considering different stages of disease. In order to evaluate which kind of genes could better represent the CTC phenotype in the two clinical settings, the contribution to CTC positivity was calculated for each of the 6 genes, at baseline and during therapy, considering all the cases who entered the study and assessable for CTC status by the EMT1 kit. In M0 cases CTC positivity rates were 5/17 (29.4%) and 2/30 (6.7%) at baseline (T0) and during therapy (any time point, Tx), respectively, whereas in M+ cases CTC positivity rates were 14/29 (48.3%) at T0 and 6/24 (25.0%) at Tx. Considering all of the samples, the two markers which mainly contributed to CTC positivity were *AKT2* and *ALDH1A*, either in M0 (4/7 CTC-positive samples, 57.1%) and M+ cases (9/20, 45.0%), followed by *MUC1* and *ERBB2* (1/7, 14.3%) in M0 cases and by *EPCAM* and *PIK3CA* in M+ cases (6/20, 30.0%), which were never detected in M0 cases, either processed with the EMT1 or EMT2 kit. Positivity rates for *MUC1* and *ERBB2* in M+ cases were 5/20 (25.0%) and 4/20 (20.0%), respectively. *TWIST1* was never detected in the overall case series, including samples processed with the EMT2 kit.

Table 4.4.2. CTC positivity rates by EMT1 and EMT2 AdnaTest kits according to the expression of different panels of genes

Clinical stage	N	Epithelial/tumor-associated genes		Epithelial/tumor-associated + EMT-related genes		Epithelial/tumor-associated + EMT-related + stemness-related genes	
		EMT1+ (%)	EMT2+ (%)	EMT1+ (%)	EMT2+ (%)	EMT1+ (%)	EMT2+ (%)
M0	21	0	1 (5)	1 (5)	6 (29)	2 (10)	8 (38)
M+	9	1 (11)	5 (56)	2 (22)	6 (67)	2 (22)	6 (67)
total	30	1 (3)	6 (20)	3 (10)	11 (37)	4 (13)	14 (47)

CTC status assessed at baseline was not associated with the clinico-pathological features of the two case series (Table 4.4.3), except for HER2 status and subtype in M0

cases only, a trend toward a significantly higher CTC positivity rate was observed in HER2-positive compared to HER2-negative cases (75% vs. 15.4%, $P = 0.0525$, two-tailed Fisher's exact test) and also, consistently, in the HER2-positive subtype compared to the luminal and TN ones (100% vs 15.4 and 50%, $P = 0.0806$, χ^2 test for trend).

Regarding the correlation between CTC status and clinical outcome, in the neoadjuvant setting, 20 and 19 out of the 22 enrolled cases were assessable for relapse-free survival and pCR, respectively. Median(range) follow-up time was 4(1-5) years and relapse events were 6 (median(range) time to relapse: 2(1-4) years). CTC status at T0 was not able to predict relapse-free survival in 16 assessable cases (75% *vs.* 57% probability for CTC-positive *vs.* CTC-negative cases, respectively, $P = 0.5235$, log-rank test). According to pathological response criteria, 2 out of 19 cases (11%) reached pCR. An association with CTC status, showing a marginal trend toward significance ($P = 0.0952$, two-tailed Fisher's exact test), could be observed in the 15 assessable cases as, 10 out of 13 cases not achieving pCR were CTC-negative, although 2/2 cases with pCR were CTC-positive, thus indicating that this result is not really informative as the number of cases is low.

In the metastatic setting, follow-up and RECIST response data were available in 25 out of 32 enrolled cases (not the same cases for the two clinical outcomes). Median(range) follow-up time was 11(2-58) months and progression events were 20 (median(range) time to progression: 10(0-39) months). Twenty-three cases were assessable for CTC status by EMT1, 18 cases experienced disease progression. CTC status assessed by AdnaTest at baseline was not able to discriminate patients at high risk of progression (9% *vs.* 20% PFS probability for CTC-positive *vs.* CTC-negative cases, respectively, $P = 0.8837$, log-rank test). Twenty-two cases were assessable for CTC status and response to therapy by instrumental analysis. According to RECIST response criteria, 1 case had CR and was CTC-negative, 8 out of 15 cases who had PR were CTC-positive, 1 case with SD was CTC-negative and 3 out of 5 cases with PD were CTC-positive. CTC status assessed at T0 was not significantly associated with response to therapy assessed by instrumental analysis in 22 assessable cases, either when considering CR+PR *vs.* SD+PD or CR+PR+SD *vs.* PD groups.

Table 4.4.3. Patients' baseline characteristics and CTC status by AdnaTest EMT1 kit

variable	Neoadjuvant setting				Metastatic setting			
	N (%)	N assessable	CTC+ (%)	*P*	N (%)	N assessable	CTC+ (%)	*P*
All patients	22	17	5 (29.4)	-	32	29	14 (48.3)	-
Age (years)	22	17			32	29		
<50	8 (36.4)	6	2 (33.3)	*1.0*[a]	5 (15.6)	5	4 (80.0)	*.1686*[a]
≥50	14 (63.6)	11	3 (27.3)		27 (84.4)	24	10 (41.7)	
ER and PgR status	22	17			32	29		
Positive for either	17 (77.3)	13	2 (15.4)	*.0525*[a]	28 (87.5)	26	12 (46.2)	*.5977*[a]
Negative for both	5 (22.7)	4	3 (75.0)		4 (12.5)	3	2 (66.7)	
HER2/*neu* status	22	17			32	29		
positive	6 (27.3)	4	3 (75.0)	*.0525*[a]	5 (15.6)	5	3 (60.0)	*.6513*[a]
negative	16 (72.3)	13	2 (15.4)		27 (84.4)	24	11 (45.8)	
Subtype	22	17			32	29		
Luminal	16 (72.7)	13	2 (15.4)	*.0806*[b]	25 (78.1)	23	10 (43.5)	*.2303*[b]
HER2+	4 (18.2)	2	2 (100)		5 (15.6)	5	3 (60.0)	
Triple-negative	2 (9.1)	2	1 (50.0)		2 (6.3)	1	1 (100)	
Histotype	18	14			32	29		
Ductal	15 (83.3)	12	3 (25.0)	*.2259*[b]	21 (65.6)	19	11 (57.9)	*.5080*[b]
Lobular	2 (11.1)	1	1 (100)		8 (25.0)	7	2 (28.6)	
Ductal+Lobular	1 (5.6)	1	0		1 (3.1)	1	0	
Papillary	0	-	-		1 (3.1)	1	0	
Inflammatory	0	-	-		1 (3.1)	1	1 (100)	
Tumor grade	16	12			27	25		
G1	0	-	-	*.2206*[a]	1 (3.7)	1	0	*.8874*[b]
G2	8 (50.0)	6	1 (16.7)		14 (51.9)	13	7 (53.8)	
G3	8 (50.0)	6	3 (50.0)		12 (44.4)	11	5 (45.5)	
Ki67 (%)	19	14			20	17		
<10%	1 (5.3)	1	1 (100)	*.1638*[a]	5 (25.0)	5	3 (60.0)	*.2801*[a]
≥10%	18 (94.7)	13	4 (30.8)		15 (75.0)	12	3 (25.0)	

[a] two tailed Fisher's exact test

[b] χ^2 test for trend

CTC status assessed at T0 was neither able to predict progression in the overall case series (*i.e.*, including M0 and M+ cases), as PFS probabilities were 9% and 16% respectively for CTC-positive and CTC-negative cases (P = 0.6412, log-rank test).

CTC status during therapy was assessable for about 50% cases in both cohorts. Low numbers do not allow drawing conclusion from longitudinal analysis.

AdnaTest seems to be a valuable method to isolate and detect CTCs. Nevertheless, correlation analyses in these two observational studies revealed that the approach does not provide clinically relevant information.

4.4.2. Technical and clinical implications of CTC monitoring by ScreenCell

CTC monitoring with ScreenCell® Cyto kit was carried out on two BC case series, which were distinct from those analyzed by AdnaTest. The study started in 2014 and involved 37 patients under neoadjuvant therapy and 24 patients starting a first-line systemic therapy (1 out of 24 starting second-line therapy) for metastatic disease. CTC analysis was performed on 9 mL of blood collected at baseline (T0) and after about 3 months of systemic therapy (T1). Blood was also collected from 4 healthy donors who were all negative for the presence of atypical or suspicious cells.

Differently from the outcome of AdnaTest CTC studies, the size-based isolation approach allowed direct visualization of CTC subpopulation, *i.e.* single CTCs and CTC clusters, hereafter referred to as circulating tumor microemboli (CTM). In a preliminary study on 14 M0 and 18 M+ BC patients, CTCs were detectable at higher rate compared to AdnaTest, as 72% positivity rate at T0 and 95% at T1 were obtained in the overall case series. Surprisingly, no significant differences were observed in terms of CTC detection frequency when categorizing patients by clinical stage or grouping samples according to the treatment time point (Table 4.4.4), and similarly median CTC counts did not differ between M0 and M+ patients (8.5 *vs.* 9 at T0, 15 *vs.* 10 at T1, and 11 *vs.* 9 overall). However, overall CTM positivity rate was significantly higher in patients with early stage disease compared to metastatic women, and the difference remained significant even when considering samples

collected at T0 (79% *vs.* 28%, $P = 0.0113$, two-tailed Fisher's exact test) or at T1 (77% *vs.* 25%, $P = 0.017$, Fisher's exact test), data consistent with the higher CTM load (3 *vs.* 0, $P = 0.0056$ at baseline; 2 *vs.* 0, $P = 0.0278$ during treatment; and 2 vs. 0, $P = 0.001$ overall; two-tailed Mann-Whitney's *U* test). This data was confirmed on the overall case series including all patients enrolled from June 2014 to July 2016: at baseline CTM positivity rates were 25/37 (68%) *vs.* 6/24 (25%) cases ($P = 0.0028$, two-tailed χ^2 test with Yates' correction), and median(range) values were 2(0-20) and 0 (0-20) ($P = 0.0078$, two-tailed Mann-Whitney's *U* test), in M0 and M+ cases, respectively.

Table 4.4.4. CTC and CTM positivity rate by ScreenCell

Clinical stage	At baseline (T0)			During therapy (T1)			At any time		
	N	CTC (%)	CTM (%)	N	CTC (%)	CTM (%)	N	CTC (%)	CTM (%)
M0	14	11 (78)	11 (79)	13	12 (92)	10 (77)	27	23 (85)	21 (78)
M+	18	13 (72)	5 (28)	12	11 (92)	3 (25)	30	24 (80)	8 (27)
overall	32	24 (75)	16 (50)	25	23 (92)	13 (52)	57	47 (82)	29 (62)

In the neoadjuvant setting CTC status, either assessed as detection frequency using a positivity cut-off ≥3 CTC/9 mL, or as mean or median value, did not differ when correlated with the clinico-pathological characteristics of the case series (Table 4.4.5). The same trend was also observed for CTM, which were analyzed using a positivity cut-off ≥3 CTC/9 mL (Table 4.4.6), although they showed significant association with some clinico-pathological categories. Actually, CTM were significantly more frequent in the blood of patients having T_1 or T_2 compared to those with T_3 or T_4 tumor stage (65% *vs.* 24%, $P = 0.0202$, two-tailed Fisher's exact test), as well as in patients with HER2-negative compared to HER2-positive tumor (58% *vs.* 18%, respectively, $P = 0.0365$, two-tailed Fisher's exact test). Moreover, the

proportion of CTM when considering the overall CTC population (*i.e.*, CTC+CTM) was higher in M0 compared to M+ cases (38% *vs.* 11%).

Table 4.4.5. CTC status at baseline by ScreenCell according to patients' characteristics in the neoadjuvant setting

variable	N (%)	CTC+ (%)*	P^a	Mean±SD CTC/9 mL	Median (range) CTC/9 mL	P^b
All patients	37	25 (68)	-	6.65±8.19	3 (0-24)	-
Age (years)						
≤50	17 (46)	10 (59)	*.4821*	4.53±5.2	3 (0-17)	*.1389*
>50	20 (54)	15 (75)		8.5±8.8	6.5 (0-35)	
Tumor size						
T_1/T_2	20(54)	12 (60)	*.3193*	6.35±6.52	3 (0-24)	*.9124*
T_3/T_4	17(46)	13 (76.5)		7.0±8.82	5 (0-35)	
Lymph-nodal status						
N_0	9 (24)	8 (89)	*.2204*	10.89±11.50	7 (0-35)	*.1416*
N_1/N_2	28 (76)	17 (61)		5.28±5.38	3 (0-17)	
Histotype						
Ductal	22 (59)	13 (59)	*.2863*	7.04±8.88	3 (0-35)	*.7490*
others	15 (41)	12 (80)		6.54±5.53	5 (0-17)	
Tumor grade						
G1/G2	11 (30)	8 (73)	*.7026*	6.73±9.72	5 (0-35)	*1*
G3	21 (57)	13 (62)		6.33±6.76	3 (0-24)	
missing	5 (13)	4 (80)	-	5.5±5.29	5 (0-24)	-
ER and PgR status						
Positive for either	18 (49)	11 (61)	*.4951*	6.22±8.61	3 (0-35)	*.3681*
Negative for both	19 (51)	14 (74)		7.06±6.61	5 (0-24)	
HER2 status						
positive	11 (30)	7 (64)	*1*	5.00±6.92	3 (0-17)	*.4473*
negative	26 (70)	18 (69)		7.35±8.22	5 (0-35)	

* cut-off ≥3 CTCs/9 mL

[a] two-tailed Fisher's exact test

[b] two-tailed Mann-Withney's *U* test

Table 4.4.6. CTM status at baseline by ScreenCell according to patients' characteristics in the neoadjuvant setting

variable	N (%)	CTM+ (%)	*P*	Mean±SD CTM/9 mL	Median (range) CTM/9 mL	*P*[c]
All patients	37	17 (46)	-	3.70±4.69	2 (0-20)	-
Age (years)						
≤50	17 (46)	8 (47)	*.9003*[a]	3.12±3.20	2 (0-10)	*.9124*
>50	20 (54)	9 (45)		4.20±5.70	1.5 (0-20)	
Tumor size						
T_1/T_2	20 (54)	13 (65)	*.0202*[b]	4.35±4.63	4 (0-20)	*.1645*
T_3/T_4	17(46)	4 (24)		2.94±4.79	1 (0-15)	
Lymph-nodal status						
N_0	9 (24)	2 (22)	*.1371*[b]	2.33±4.47	0 (0-12)	*.9681*
N_1/N_2	28 (76)	15 (54)		4.14±4.75	3 (0-20)	
Histotype						
Ductal	22 (59)	9 (41)	*.6828*[a]	3.55±5.11	1 (0-20)	*.4839*
others	15 (41)	8 (53)		3.93±4.15	3 (0-15)	
Tumor grade						
G1/G2	11 (30)	5 (45)	*1*[b]	4.0±5.42	1 (0-15)	*1*
G3	21 (57)	9 (43)		3.40±4.66	1 (0-20)	
missing	5 (13)	3 (60)	-	4.40±3.78	5 (0-10)	-
ER and PgR status						
Positive for either	18 (49)	10 (56)	*.4170*[a]	4.72±5.77	3 (0-20)	*.4009*
Negative for both	19 (51)	7 (37)		2.74±3.25	1 (0-10)	
HER2 status						
positive	11 (30)	2 (18)	*.0365*[b]	1.45±2.61	0 (0-8)	*.0324*
negative	26 (70)	15 (58)		4.70±5.10	3.5 (0-20)	

* cut-off ≥3 CTCs/9 mL

[a] two tailed χ^2 test with Yates' correction

[b] two-tailed Fisher's exact test

[c] two-tailed Mann-Withney's *U* test

Correlation analysis between CTC or CTM status and the response to therapy assessed by histological analysis were performed. Eight out of 36 assessable cases reached pathologic complete response (pCR). A trend toward CTC and CTM increase, either in terms

of detection frequency or number, was observed when categorizing patients according to the extent of residual tumor after neoadjuvant therapy (Table 4.4.7). At T0 CTC positivity rates were 50%, 69% and 80%, respectively in cases classified as complete responders (CR), partial responders (PR) or with stable disease (SD), whereas values at T1 showed only slight differences among the three categories (around 90% when considering the overall case series). In a similar manner, CTM frequency at T0 increased from ~40% in cases with CR or PR to 60% in cases with SD, whereas no differences were observed during therapy. When grouping patients with PR or SD in a unique category (non-responders), at baseline higher CTC detection rate was observed in non-responders compared to patients with pCR (71% *vs.* 50%, n.s.), whereas opposite trend was obtained at T1. On the contrary, CTM frequencies were comparable at T0 in responders and non-responders (46% *vs.* 54%, respectively) but, similarly to CTCs, at T1 CTM were more frequently detected in non-responders (65% *vs.* 35%, respectively, n.s.). At T1 all categories showed an increase in CTC load compared to baseline levels. In particular, statistically significant increase in median CTCs was reached in CR (4.7-fold, $P = 0.044$, two-tailed Mann-Whitney's *U* test) and PR (2.3-fold, $P = 0.035$, two-tailed Mann-Whitney's *U* test) groups.

In the M+ setting statistical analyses were focused on the correlation between the CTC (Table 4.4.8) or CTM (Table 4.4.9) status, assessed at T0 and T1, and the therapy response evaluated by imaging after about 3 months from the start of treatment. Data were available for 21 out of 24 cases at T0 and 18 out of 19 matched samples collected at T1. None patient reached CR, 11 showed PR and 10 were classified as non-responders, 7/10 classified as SD and 3/10 experiencing progression (PD). Overall, CTC and CTM status assessed at T0 and T1, either in terms of detection frequency or number, was not informative on therapy response evaluated by instrumental analysis.

Representative images of CTCs and CTM from BC patients are reported in Figure 4.4.2.

Table 4.4.7. CTC and CTM status at baseline and during therapy in M0 cases according to pathologic response

pathologic response	At baseline					During therapy				
	N	CTC+ (%)	CTC mean±SD median (range)	CTM+ (%)	CTM mean±SD median (range)	N	CTC+ (%)	CTC mean±SD median (range)	CTM+ (%)	CTM mean±SD median (range)
CR	8	4 (50)	4.56±5.13 3 (0-13)	3 (37.5)	2.78±2.99 3 (0-8)	7	7 (100)	24.57±33.61 14 (1-97)	3 (43)	14.71±36.31 0 (0-97)
PR	23	16 (69)	5.78±6.14 3 (0-24)	10 (43)	3.87±5.33 1 (0-20)	19	17 (89)	17.79±24.40 7 (0-87)	6 (32)	5.0±10.21 1 (0-44)
SD	5	4 (80)	14.4±12.82 12 (0-35)	3 (60)	4.6±4.56 4 (0-12)	5	4 (80)	15.80±14.15 18 (0-35)	2 (40)	1.6±2.30 0 (0-5)

cut-off ≥3 CTCs/9 mL

Table 4.4.8. CTC status at baseline and during therapy in M+ cases according to instrumental response									
timing	N	overall		Response to therapy by imaging					
				Partial response		Stable disease		Progressive disease	
		CTC+ (%)	CTC mean±SD	CTC+ (%)	CTC mean±SD	CTC+ (%)	CTC mean±SD	CTC+ (%)	CTC mean±SD
T0	24	17/24 (71)	11.26±12.47	8/11 (73)	13.37±14.38	6/7 (86)	5.86±3.08	2/3 (67)	17.67±11.93
T1	19	12/19 (63)	6.57±5.28	5/10 (50)	6.7±5.87	5/5 (100)	5.2±1.30	2/3 (67)	8.67±9.02
cut-off ≥5 CTCs/9 mL									

Table 4.4.9. CTM status at baseline and during therapy in M+ cases according to instrumental response									
timing	N	overall		Response to therapy by imaging					
				Partial response		Stable disease		Progressive disease	
		CTM+ (%)	CTM mean±SD	CTM+ (%)	CTM mean±SD	CTM+ (%)	CTM mean±SD	CTM+ (%)	CTM mean±SD
T0	24	6/24 (25)	2.17±5.31	3/11 (27)	3.27±7.13	2/7 (29)	1.71±4.11	1/3 (33)	0.67±1.15
T1	19	7/19 (37)	2.16±5.63	6/10 (60)	4.0±7.44	0/15 (0)	0	1/3 (33)	0.33±0.58
cut-off ≥1 CTM/9 mL									

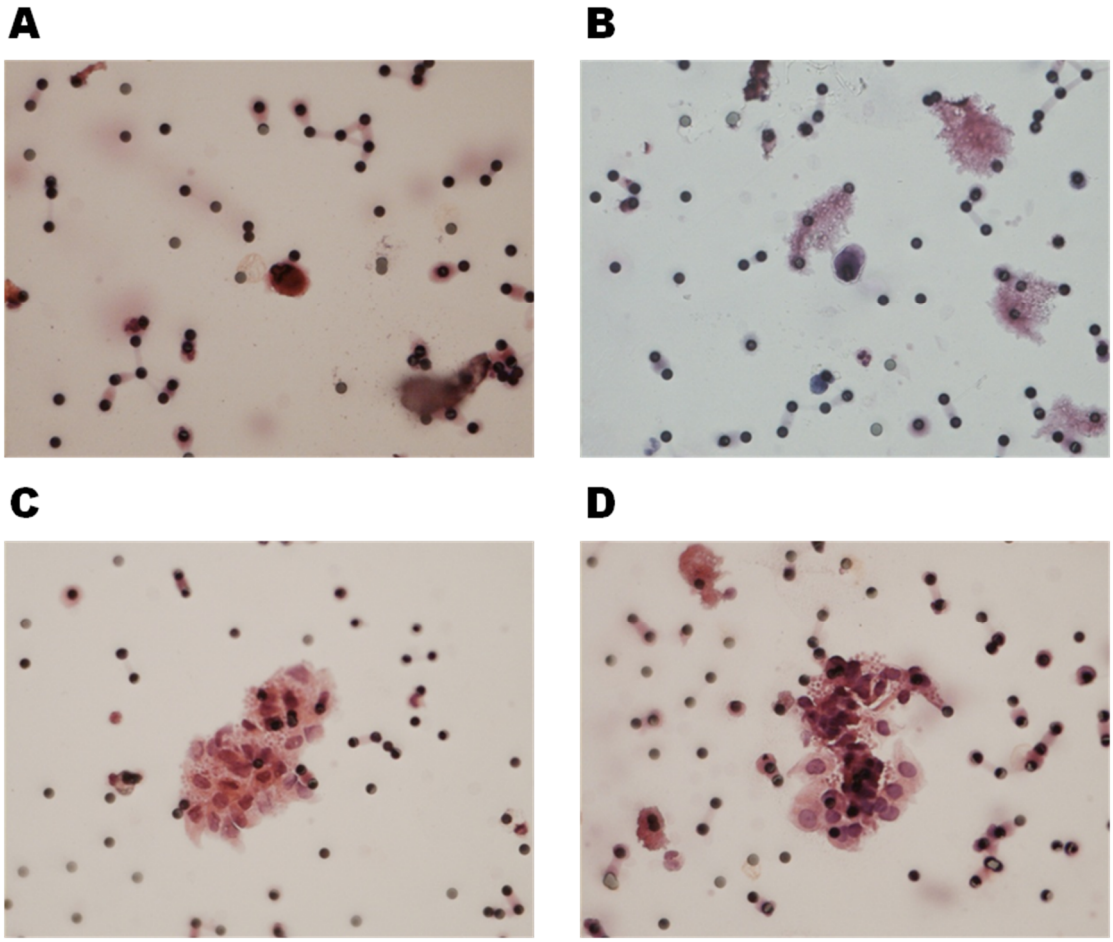


Figure 4.4.2. Single CTCs and CTM in M0 and M+ breast cancer patients. Images (400X total magnification, hematoxylin and eosin staining) depict a single CTC in a cytological blood sample from patients (**A**) under neoadjuvant therapy for M0 BC or (**B**) with M+ BC before starting first-line therapy, and a CTM mixed with platelets in a cytological blood sample from patients with (**C**) M0 or (**D**) M+ BC before starting therapy.

4.5. Clinical significance of CTC/metastasis-specific genes in breast cancer

4.5.1. Selection criteria of genes from experimental metastatic models for CTC-clinical studies

Metastatic experimental models may represent a powerful source of information to better understand the biology of the metastatic process and to find new metastasis-related and prognostic/predictive biomarkers. Genes identified in experimental models as CTC-specific, in particular, deserve gene expression analysis not only in PT but also in CTC samples of neoplastic patients, in addition to the conventional evaluation of the CTC status (*i.e.*, based on the expression of epithelial/tumor-associated genes in CTC-enriched blood samples) or as an alternative to the direct CTC count. To this aim, the association between such genes identified in the MDA-MB-231 CTC-model and the clinico-pathological features of BC case series were investigated.

Twelve out of 192 genes up-regulated in CTCs compared to solid lesions in GEP1 and GEP2 were selected for clinical studies, according to their biological role, with special reference to the metastatic process, as reported in literature (Table 4.5.1). Moreover, some of the genes up-regulated in CTC *vs.* PT in GEP1 experiment and selected for functional validation studies were also included (Table 4.5.1). Down-regulated genes were excluded from the panel of candidate genes for clinical studies since they were supposed to be less frequently detectable and less informative compared to CTC-specific up-regulated genes.

Further genes known to be involved in lung colonization, stemness and resistance to therapy, and in some cases also proved to have clinical significance in BCs, were added to the panel of candidate genes for CTC monitoring studies, in an attempt to investigate whether they can provide clinically useful information also when detected in the 'liquid biopsy' (Table 4.5.1).

Finally, agreement or discrepancy in ER expression between the PT and CTCs were also assessed.

Table 4.5.1. Panel of 22 genes for clinical studies

CTC-specific	Up-regulated in CTC *vs.* PT	Metastasis and/or stemness-related	Classical predictive marker
ADPRHL1	**ELF3**	**OPN (SPP1)**	**ESR1**
CRIP1	**TFF1**	**PTPN11**	
FCF1	**TFF2**	**SOX2**	
FIS1	**TFF3**	**SPARC**	
GAS2L1		**YAP1**	
GIGYF1			
HDAC10			
KLC2			
NR4A1			
STRN4			
TAF6			
TPPP			

4.5.2. CTC/metastasis-specific genes from experimental models are detectable in CTCs from clinical samples

Data from a GEP experiment with the DASL assay were preliminarily analyzed in order to explore the detection rate of the 22 selected genes in blood samples from 13 M+ patients (GEP of 7/13 cases reported in Fina et al., 2015), collected before starting primary systemic therapy, and from 2 female healthy donors, processed with the EMT1 kit.

Overall, the majority of the 21/22 genes assessable by the DASL assay (*TFF2* was not detectable) was detectable in patients at higher frequency compared to healthy donors (mean detection rate 52% *vs.* 31%, respectively) and, if considering genes detected in both groups, their signal intensities were overall higher in patients compared to healthy donors (Figure 4.5.1).

Gene expression analysis was then performed on residual cDNA samples using a low-density array. Samples which did not pass the quality control criteria were excluded and

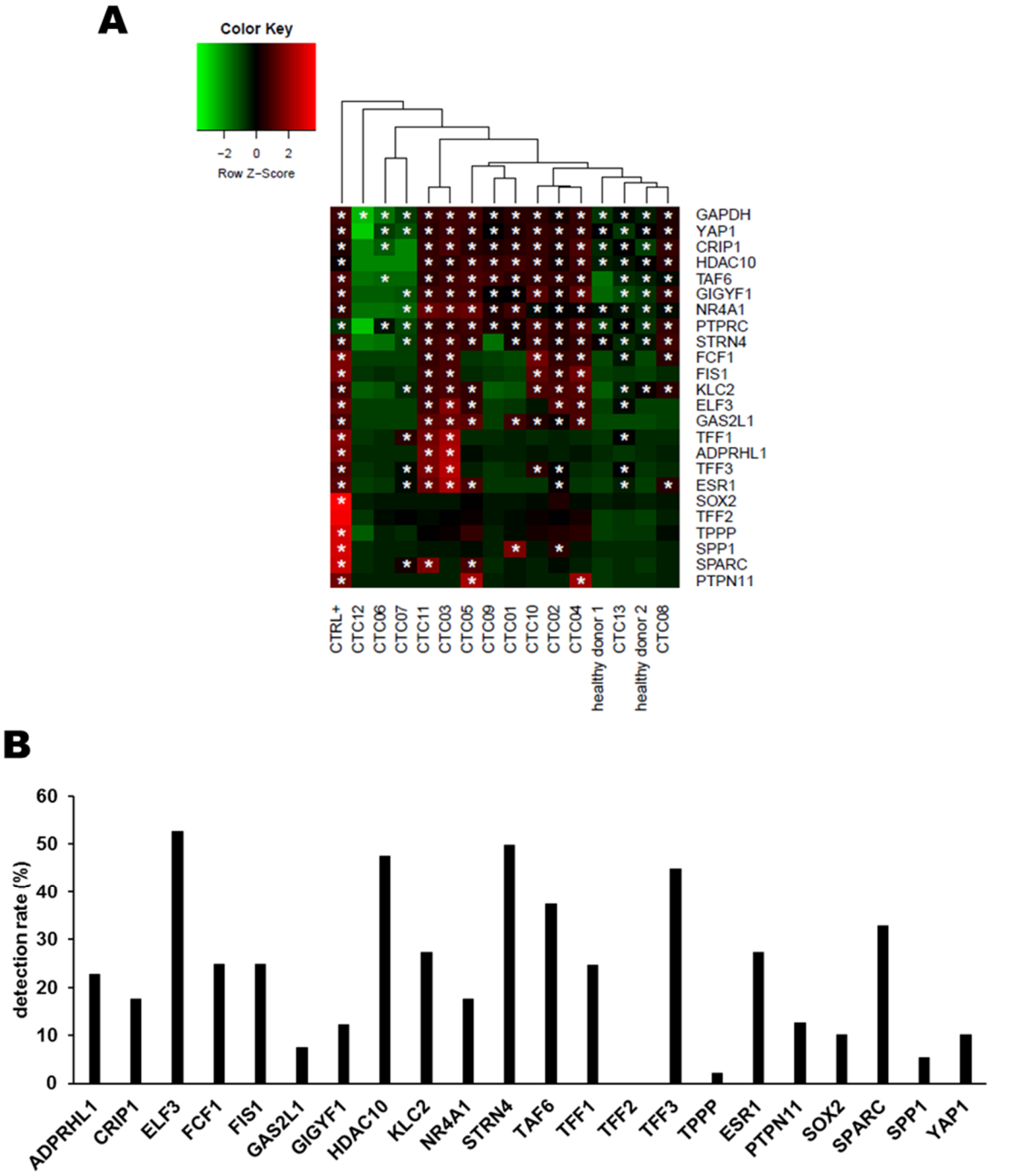


Figure 4.5.1. Detection rate of CTC/metastasis-specific genes in CTCs from BC patients. Heat map (**A**) represents the signal intensity (Row Z-score) and the detection status (detected if detection p-value <0.01) for the 22-gene panel and two control genes (*GAPDH* and *PTPRC*) in a positive control sample (UHR) and blood samples from 13 BC patients and 2 healthy donors processed with the EMT1 AdnaTest kit and profiled by the WG HT DASL assay. Bar chart (**B**) represents the detection rate (percentage) of each gene on a low-density array in blood samples of 49 BC patients processed with the EMT1 AdnaTest kit before starting therapy.

positivity for each of the 22 genes was assigned according to the threshold values previously set on a case series of 6 female healthy donors (Figure 4.5.2). The lowest mean equivalent C_q value for each gene was adopted as positivity cut-off in the analysis of patients' samples.

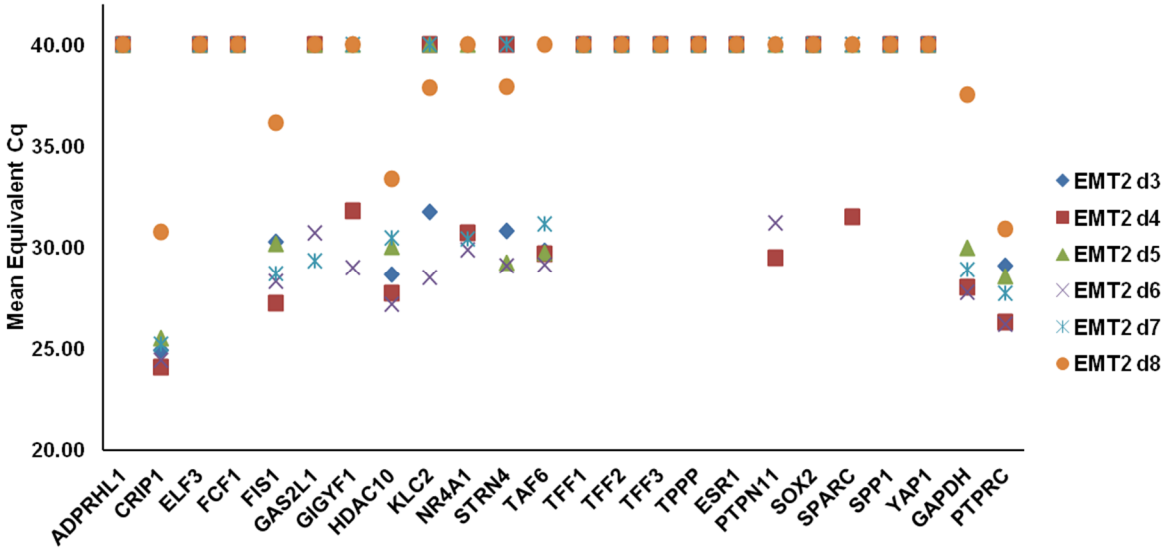


Figure 4.5.2. Expression level of CTC/metastasis-specific genes in healthy donors. Dot plot represents the mean equivalent C_q of 22 CTC/metastasis-specific genes and the two control genes *GAPDH* and *PTPRC* in a series of blood samples from 6 female healthy donors processed by EMT2 AdnaTest kit.

Comparative analysis (Figure 4.5.3) was performed on 11 pairs of CTC samples processed in parallel with EMT1 and EMT2 kits, (5 belonging to the M0 setting and 6 to the M+ setting, 6 out of 11 collected at baseline). Detection rate was also evaluated on 49 cases (20 of the M0 and 29 of the M+ setting), analyzed for CTC status by EMT1 kit before starting primary systemic therapy (T0). Molecular profile analysis of CTCs collected after the first cycle of therapy (T1) was performed on 30 cases. *TFF2* was never detected. Concordance in detection frequencies between samples processed by EMT1 and matched samples processed by EMT2 was poor in the majority of cases, except for *TFF3* and *TFF1* (Cohen's $\kappa = 0.744$, 95%CI (0.3-1.0)), *FCF1* and *SPARC* (Cohen's $\kappa = 0.633$, 95%CI (0.2-1.0)), all showing good agreement, and *NR4A1* (Cohen's $\kappa = 0.421$, 95%CI (-0.034-0.877)), showing a moderate agreement. The majority of detected genes was more frequently expressed in samples processed by the EMT2 kit, whereas EMT1 outperformed EMT2 only for *ADPRHL1*

gene. *SPARC*, *TPPP*, *TAF6*, *STRN4*, *HDAC10* and *FCF1* showed the same positivity rate both in EMT1 and EMT2 samples.

Genes with the highest detection rate in the whole case series of CTCs collected at T0 and processed by EMT1 were *ELF3* (21/49, 43%), *STRN4* (20/49, 41%), *HDAC10* (19/49, 39%) and *TFF3* (18/49, 37%).

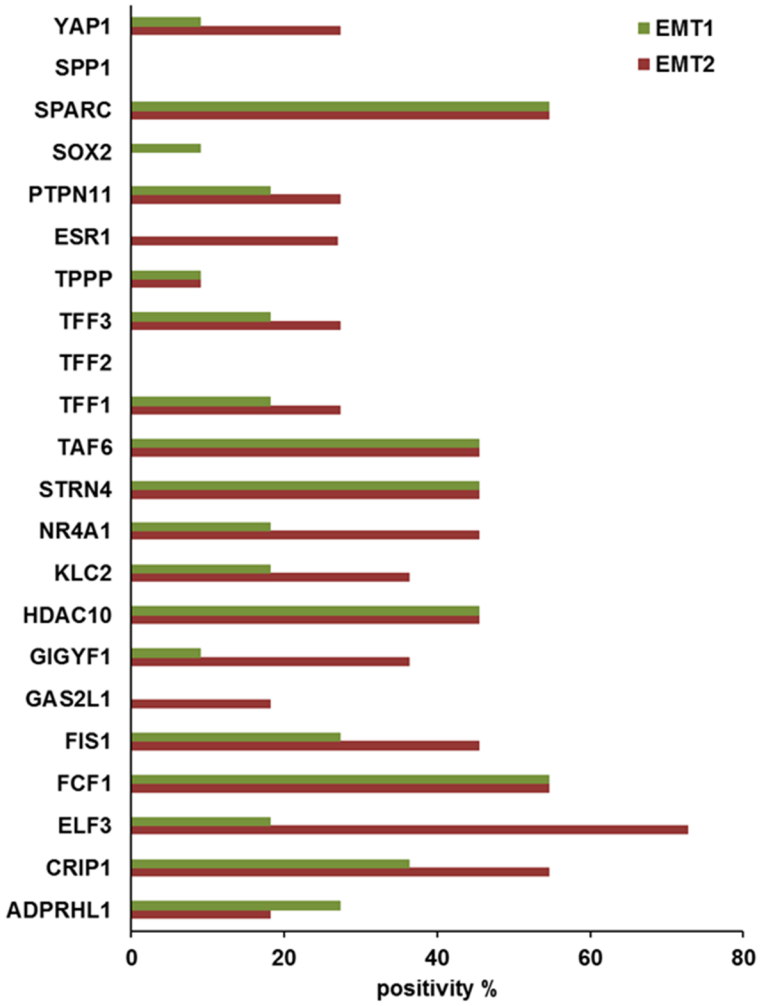


Figure 4.5.3. Detection rate of CTC/metastasis-specific genes in matched CTC samples isolated by AdnaTest EMT1 and EMT2 kits. Bar chart reports the percentage frequency of detection of each of the 22 genes in matched blood samples processed by AdnaTest EMT1 and EMT2 kits and isolated from 11 BC patients at different time points.

Interestingly, concordance between CTC positivity assessed by AdnaTest (AT) and the 22-gene panel (GP) was poor (Cohen's $\kappa = 0.117$, 95%CI (-0.046-0.280)), as 21 out of 26 AT-negative samples were rated as CTC-positive according to the detection thresholds of the

GP (Table 4.5.2). Overall, GP outperformed AdnaTest as roughly 2-fold higher CTC positivity rate was obtained if compared with CTC status assessment by using *EPCAM*, *MUC1*, *ERBB2*, *PIK3CA*, *TWIST1* and *ALDH1* only (86% *vs.* 41% for GP and AT, respectively).

Table 4.5.2. Concordance between AdnaTest and the 22-gene panel

N	AT+/GP+ (%)	AT-/GP- (%)	AT+/GP- (%)	AT-/GP+ (%)
44	17 (38.6)	5 (11.4)	1 (2.3)	21 (47.7)

4.5.3. CTC/metastasis-specific genes detected in CTCs predict disease progression

The expression rate of the 21 detectable genes was evaluated as a function of clinical setting (20 M0 *vs.* 29 M+ cases) and BC subtype (38 luminal *vs.* 8 HER2-positive *vs.* 3 TN). The positivity frequency for each gene was calculated on the number of positive samples. In this case CTC positivity was re-assigned considering positive those samples with at least one positive AT or GP marker. On the basis of these criteria, overall positive cases at T0 were 15/20 M0 and 25/29 M+, while positive cases according to the subtype were 32/38 luminal, 5/8 HER2-positive and 3/3 TN. As reported in Figure 4.5.4, some differences in detection rates according to the clinical setting could be observed, as 11 genes were more frequently detected in M0 cases and the remaining were more frequently detected in M+ cases. In particular, the most marked differences between M0 and M+ were observed for *TFF1* (0 *vs*. 39%, $P = 0.0063$, two-tailed Fisher's exact test) and *TFF3* (13% *vs.* 64%, $P = 0.0028$, two-tailed Fisher's exact test).

No statistically significant differences could be identified among BC subtypes, except for *GIGYF1* ($P = 0.0097$, χ^2 test for trend), which was detectable in 3% only of luminal cases and in 38% and 33% of HER2-positive and TN cases, respectively.

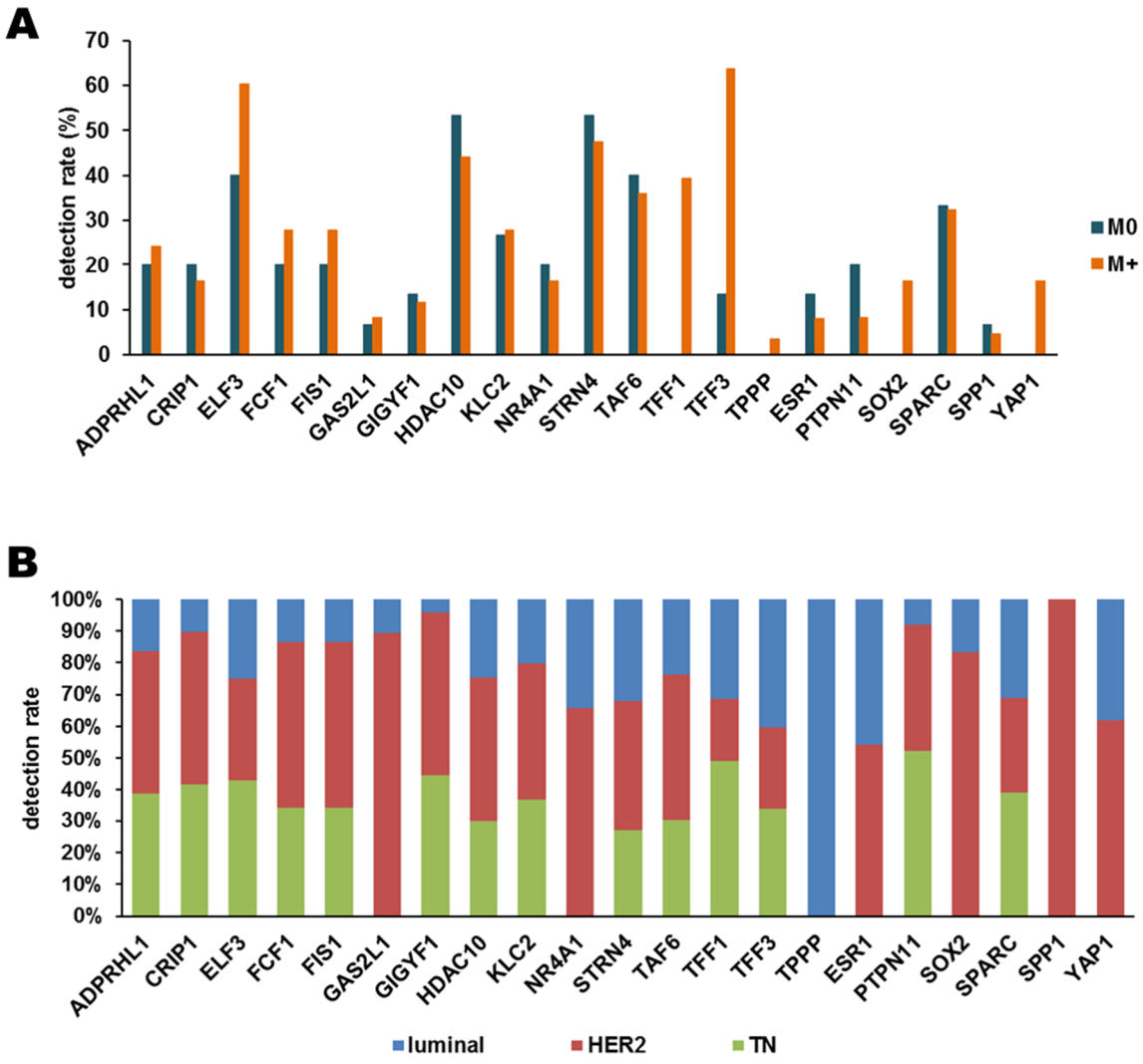


Figure 4.5.4. Positivity frequency of the 22-gene panel according to the clinical setting and BC subtypes. Bar chart reports the detection rate of each CTC/metastasis-specific gene expressed as percentage of the number of cases with positive gene status on the number of cases positive for at least one of the markers assessed by AdnaTest or the 22-gene panel, according to the (**A**) clinical setting and (**B**) the BC subtype.

The expression status of each gene was correlated with progression-free survival probability. Follow-up data were available in 41 out of 49 cases, with 25 unfavorable events. As reported in Table 4.5.3, the analysis revealed that the status (negative or positive) of *TFF3*, *TFF1*, *YAP1* and *SOX2* genes was able to significantly discriminate between cases at higher risk of relapse or progression. *ELF3* was proven to be informative since cases with *ELF3* positive status had 2.1 higher risk to undergo disease progression (20% PFS probability compared to 46% in *ELF3* negative cases), and there was a trend in favor of a

statistical significance ($P = 0.0650$, Cox proportional-hazards regression model). Kaplan-Meier plot for *TFF3*, *TFF1*, *ELF3*, *SOX2* and *YAP1* are represented in Figure 4.5.5.

The clinical significance of the expression level of such genes in primary BCs was evaluated in publicly available datasets of untreated and node negative BC patients in correlation with distant metastases-free survival. Interestingly, *ELF3*, when expressed at high level, and *TFF1* and *YAP1*, when expressed at low level, predicted poor clinical outcome, whereas the expression levels of *TFF3* and *SOX2* were not informative when assessed at the PT level (Figure 4.5.6; statistical significance was assessed by log-rank test).

Bivariate analysis was carried out in order to evaluate whether each CTC-specific clinically informative gene maintained its predictive role even in the presence of information provided by the other covariates (including stage of disease, CTC status assessed by AdnaTest, and the expression of the other investigated genes), using “M0”, “negative CTC status” and “negative gene status” as reference categories (Tables 4.5.4-8; statistical significance was assessed by Cox proportional-hazards regression model). The outcome of such an analysis, although providing interesting hints, should be considered with caution, due to the limited number of cases within subsets.

The prognostic significance was maintained independently from AdnaTest outcome for all of the investigated genes, with minimal variations of the unadjusted hazard ratio (HR) observed in the univariate analysis (Table 4.5.3), except for *ELF3*, which showed only a trend toward significance. However, all genes failed to predict progression-free survival (PFS) when adjusted for the clinical setting; only *TFF1* expression was suggestive, although not significantly, of disease progression.

The bivariate analysis carried out to investigate whether each of the 5 genes maintained its prognostic relevance even in the presence of information provided by the other genes showed that *ELF3* and *SOX2* failed to independently predict disease progression. *YAP1* maintained an independent predictive role in the presence of information provided by *ELF3*. Conversely, *TFF1* and *TFF3* were significantly predictive of progression-free survival when adjusted for the other investigated genes, but a significant

interaction (*P* = 0.026, Cox proportional-hazards regression model) was present between them and affected their predictive ability. In fact, *TFF3* failed to significantly influence PFS probability in the presence of *TFF1*, which conversely maintained its prognostic role only within tumors with *TFF3*-negative CTCs.

Clustering analysis according to the normalized expression of the 22-gene panel in blood samples collected before starting and during therapy, allowed identifying some clusters of genes which showed high expression in different groups of patients (Figure 4.5.7). At T0 (Figure 4.5.7, Panel A), gene cluster "G1" included 9 genes (*HDAC10*, *FIS1*, *STRN4*, *CRIP1*, *TAF6*, *KLC2*, *FCF1*, *GIGYF1* and *PTPN11*) which showed high expression in 5 cases, 3 M0 (1 lost at follow-up and 2 without evidence of disease) and 2 M+ patients who experienced progression. Interestingly, gene cluster "G2" included 5 genes (*NR4A1*, *ELF3*, *SPARC*, *GAS2L1* and *TFF3*) which showed high expression in a group of 6 relapsed patients, 5 M+ and 1 M0. Finally, gene cluster "G3" included the remaining 7 genes (*ADPRHL1*, *SPP1*, *ESR1*, *TFF1*, *SOX2*, *YAP1* and *TPPP*) which were highly expressed in 13 patients, 7 relapsed out of 8 M+ and 3 relapsed out of 5 M0 (follow-up data were available for 10 cases only). At T1, one cluster including 10 genes (*GIGYF1*, *ADPRHL1*, *ELF3*, *TFF3*, *TFF1*, *SPP1*, *ESR1*, *TPPP*, *YAP1* and *SOX2*) was clearly identifiable; such genes were highly expressed in 6 M+ (5 relapsed) and 6 M0 (3 relapsed, 1 lost at follow-up) patients (Figure 4.5.7, Panel B).

4.5.4. CTCs detected in ER-positive BCs are prevalently ER-negative

Concordance in ER status between CTC and PT was assessable in 40 cases. High discrepancy was observed (29/40 cases, 72.5%) as only 4 cases had concordant ER positive status at PT and CTC level, whereas 7 cases showed negative ER status for both tumor sites. All of the other patients had ER-positive BCs but did not express *ESR1* at CTC level.

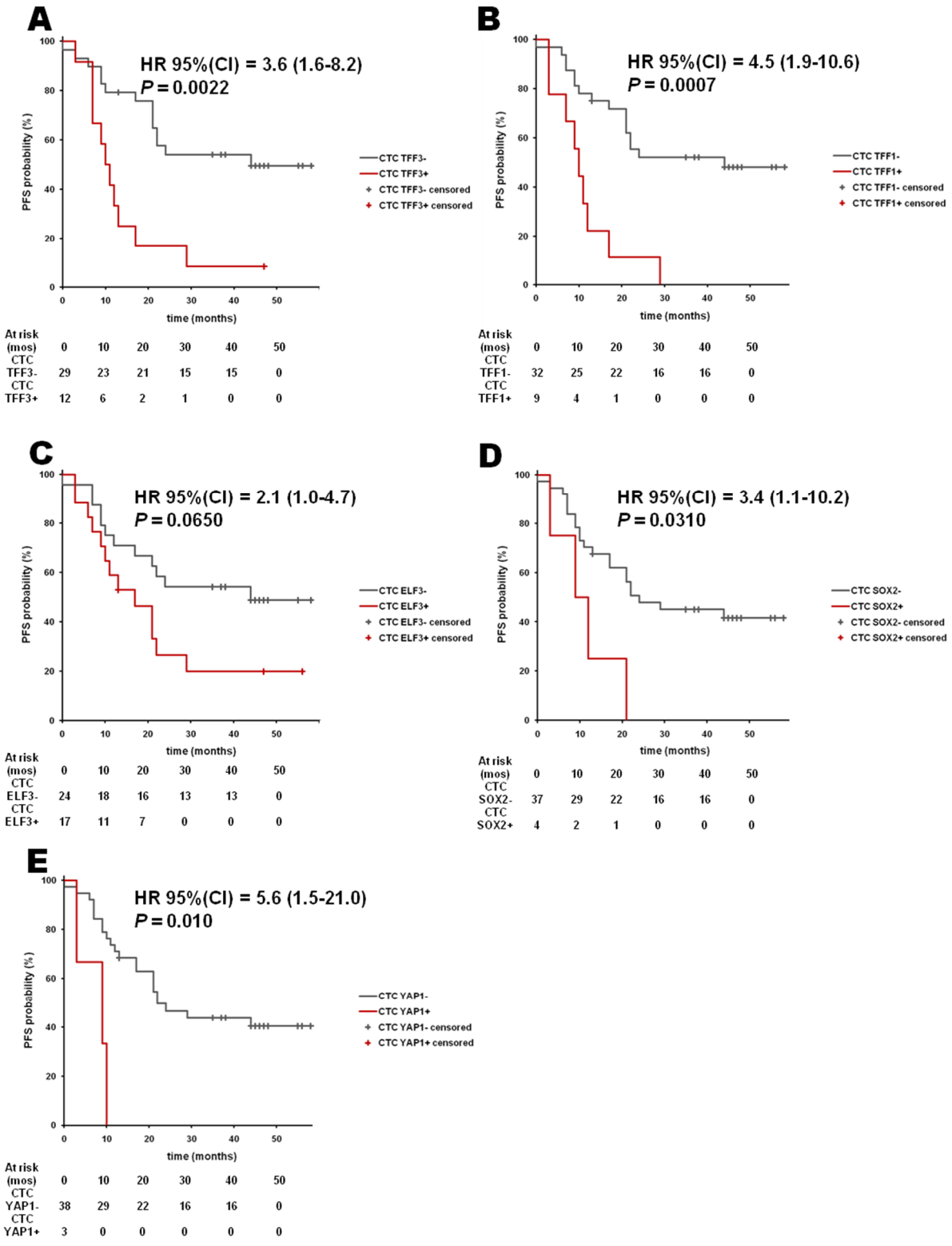


Figure 4.5.5. Progression-free survival probability according to the status of CTC/metastasis-specific genes detected in CTCs from BC patients. Survival analysis by Cox regression model and Kaplan-Meier plots of progression-free survival according to the status (negative or positive) of (**A**) *TFF3*, (**B**) *TFF1*, (**C**) *ELF3*, (**D**) *SOX2* and (**E**) *YAP1* CTC/metastasis-specific genes, assessed in CTC samples of 41 BC cases before starting therapy.

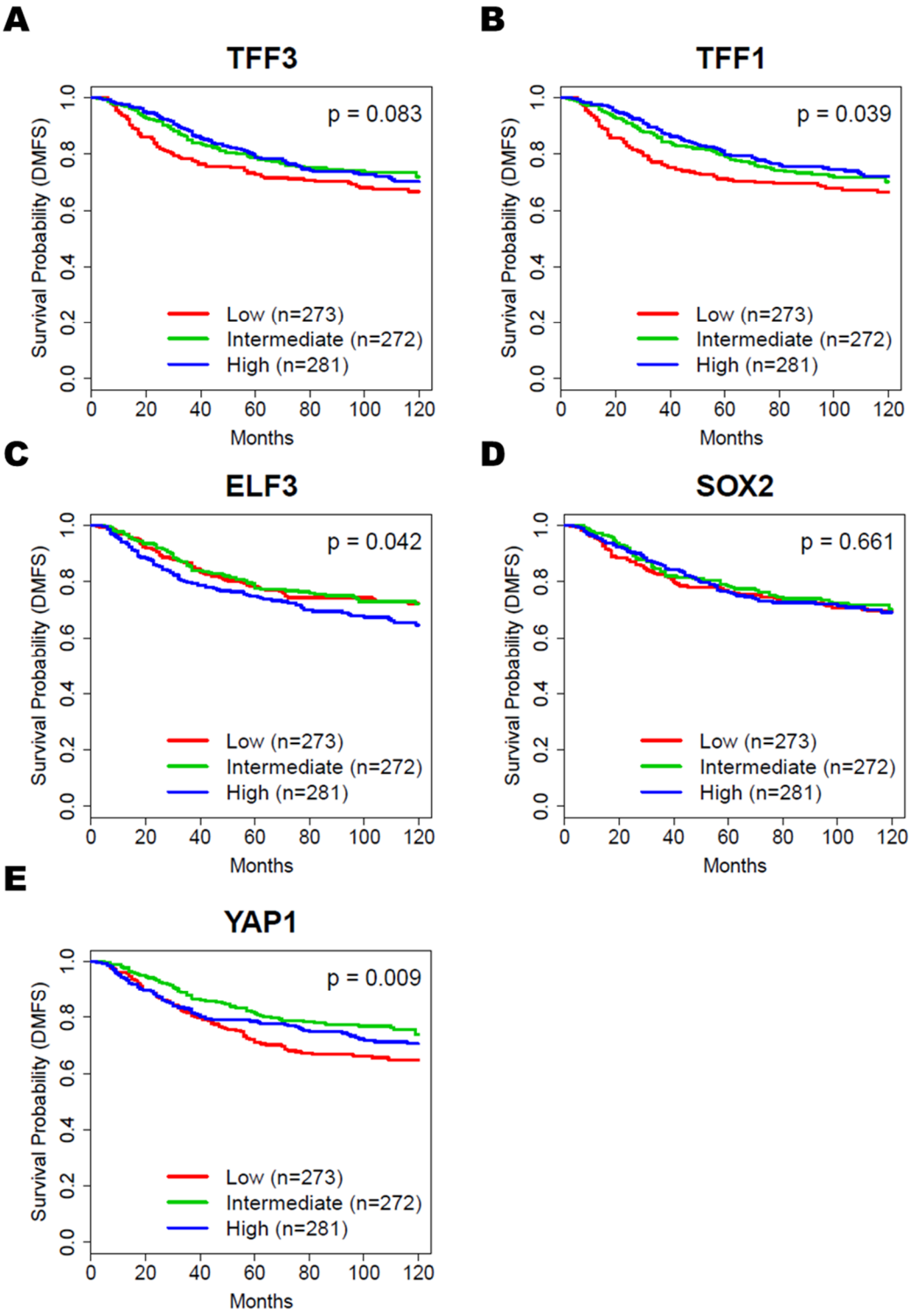


Figure 4.5.6. Distant metastasis-free survival probability according to the expression data of CTC/metastasis-specific genes in primary BCs. Kaplan-Meier plots of distant metastasis-free survival probability according to the expression level of (**A**) *TFF3*, (**B**) *TFF1*, (**C**) *ELF3*, (**D**) *SOX2* and (**E**) *YAP1* CTC/metastasis-specific genes, assessed in publicly available datasets from untreated and LN negative primary BCs.

Table 4.5.3. Progression-free survival probability according to the positivity by AdnaTest and by the 22-gene panel

Genes	Negative		Positive				
	N	PFS probability (%)	N	PFS probability (%)	HR	95% CI	*P***
CTC status by AdnaTest	20	42	16	25	1.0	0.6-1.6	*.8608*
ADPRHL1	30	34	6	33	1.6	0.5-4.6	*.4054*
CRIP1	30	31	6	50	0.6	0.2-2.1	*.4724*
ELF3	19	46	17	20	2.1	1.0-4.7	*.0650*
FCF1	27	35	9	33	1.3	0.6-3.0	*.4701*
FIS1	32	39	9	33	1.1	0.4-2.7	*.9256*
GAS2L1	33	34	3	33	1.2	0.3-5.2	*.7814*
GIGYF1	31	30	5	60	0.5	0.1-2.0	*.3144*
HDAC10	22	33	14	36	1.0	0.4-2.2	*.9405*
KLC2	32	38	9	33	1.1	0.5-2.8	*.7948*
NR4A1	30	38	6	17	1.6	0.6-4.4	*.3234*
STRN4	21	34	15	33	1.3	0.6-2.9	*.5240*
TAF6	28	37	13	38	1.0	0.4-2.1	*.8206*
TFF1	27	46	9	0	4.5	1.9-10.6	*.0007*
TFF3	24	47	12	8	3.6	1.6-8.2	*.0022*
TPPP*	35	35	1	0	-	-	-
ESR1	33	34	3	33	1.5	0.3-6.4	*.5880*
PTPN11	32	32	4	50	0.7	0.2-2.8	*.5729*
SOX2	32	39	4	0	3.4	1.1-10.2	*.0310*
SPARC	25	37	11	27	1.3	0.6-3.1	*.5024*
SPP1*	34	30	2	100	-	-	-
YAP1	33	38	3	0	5.6	1.5-21.0	*.0100*

* Genes with ≤2 CTC+ cases were not analyzed

** Cox proportional-hazards regression model

Table 4.5.4. Bivariate analysis of 4-year progression-free survival as a function of *ELF3* status (N = 41)

variable	HR	95% CI	*P*	HR for ELF3 status	95% CI	*P*
Clinical setting	4.3	1.7-11.8	*.0029*	1.3	0.6-3.0	*.5099*
CTC status by AdnaTest	0.9	0.5-1.7	*.7462*	2.1	1.0-4.8	*.0626*
***TFF1* status**	3.8	1.4-10.3	*.0078*	1.3	0.5-3.3	*.5492*
***TFF3* status**	3.3	1.2-9.3	*.0241*	1.1	0.4-3.1	*.7895*
***SOX2* status**	2.7	0.9-8.4	*.0861*	1.9	0.8-4.2	*.1349*
***YAP1* status**	4.1	1.0-16.1	*.0475*	1.8	0.8-4.1	*.1820*

Table 4.5.5. Bivariate analysis of 4-year progression-free survival as a function of *TFF1* status (N = 41)

variable	HR	95% CI	*P*	HR for *TFF1* status	95% CI	*P*
Clinical setting	3.7	1.3-10.2	*.0129*	2.3	0.9-5.9	*.0745*
CTC status by AdnaTest	1.0	0.5-1.7	*.8881*	4.5	1.9-10.6	*.0007*
TFF3* status	1.9	0.5-7.7	*.3934*	2.6	0.6-11.4	*.2043*
***SOX2* status**	1.9	0.6-6.3	*.2881*	3.9	1.6-9.9	*.0040*
***YAP1* status**	2.1	0.5-8.8	*.3221*	3.8	1.4-1.0	*.0072*

*interaction

Table 4.5.6. Bivariate analysis of 4-year progression-free survival as a function of *TFF3* status (N = 41)

variable	HR	95% CI	*P*	HR for TFF3 status	95% CI	*P*
Clinical setting	3.7	1.3-10.5	*.0133*	1.9	0.8-4.6	*.1707*
CTC status by AdnaTest	0.9	0.5-1.7	*.8492*	3.6	1.6-8.2	*.0022*
***SOX2* status**	2.5	0.8-7.7	*.1206*	3.3	1.4-7.7	*.0059*
***YAP1* status**	2.7	0.7-10.8	*.1657*	3.1	1.3-7.5	*.0131*

Table 4.5.7. Bivariate analysis of 4-year progression-free survival as a function of *SOX2* status (N = 41)

variable	HR	95% CI	*P*	HR for SOX2 status	95% CI	*P*
Clinical setting	4.5	1.7-11.7	*.0023*	1.8	0.6-5.7	*.2920*
CTC status by AdnaTest	1.0	0.5-1.7	*.8491*	3.4	1.1-10.2	*.0313*
***YAP1* status**	4.0	1.0-16.7	*.0589*	2.4	0.7-8.2	*.1564*

Table 4.5.8. Bivariate analysis of 4-year progression-free survival as a function of *YAP1* status (N = 41)

variable	HR	95% CI	*P*	HR for YAP1 status	95% CI	*P*
Clinical setting	4.5	1.7-11.7	*.0021*	3.3	0.9-12.5	*.0762*
CTC status by AdnaTest	0.9	0.5-1.6	*.6608*	5.9	1.6-22.2	*.0092*

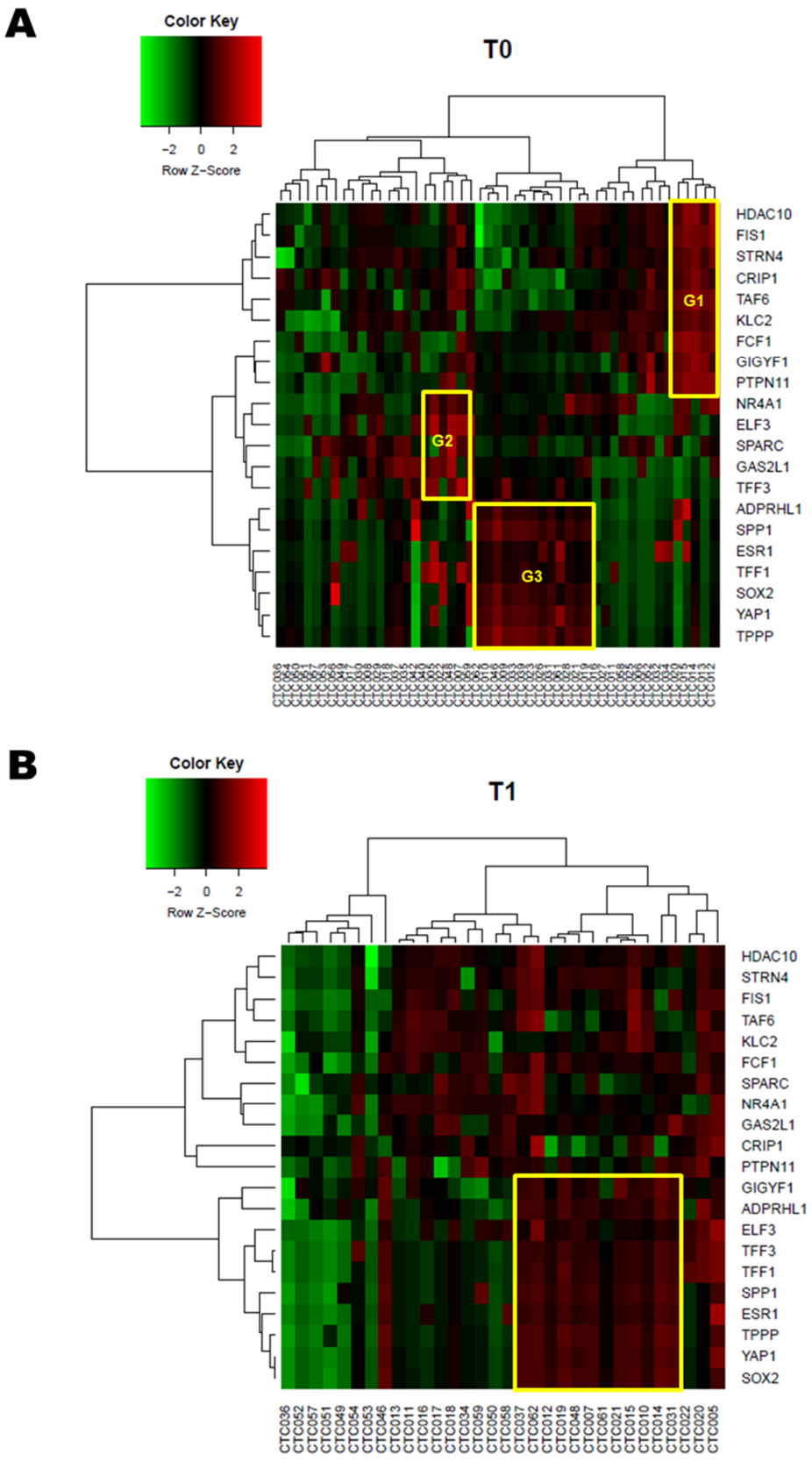


Figure 4.5.7. Clustering analysis of CTC samples at baseline and during therapy according to the expression of 22 genes. Heat maps represent the result of clustering analysis in (**A**) 49 CTC samples collected at baseline and (**B**) 32 CTC samples collected during treatment, according to the normalized expression of 22 genes.

5. DISCUSSION

Tumor cell dissemination is definitely a hallmark of cancer, but dynamics and molecular mechanisms required for the successful completion of the metastatic process are not fully understood, and at present metastasis onset represents the main cause of BC-related deaths. In an attempt to contribute to the understanding of the metastatic dissemination, this dissertation work explored the hypothesis that by comparing the molecular profile of CTCs with those of the solid lesions in BC experimental models new metastasis-associated and clinically useful biomarkers can be identified.

Preliminary studies were therefore designed to model BC metastases to provide information for the first time on the hematogenous dissemination potential of four BC cell lines belonging to different molecular subtypes. Consistently with the growth rate of the PTs, mouse xenograft models generated using the HER2-positive BT-474 and MDA-MB-453 and the basal A MDA-MB-468 cell lines showed lower CTC load compared to CTCs generated by the highly proliferating and metastatic MDA-MB-231 cell line, belonging to the more aggressive basal B molecular subtype. Importantly, similarly to what obtained with the MDA-MB-231 model, we observed pleomorphism in the CTC population, since we were able to detect both single CTCs and CTC clusters, also in the so called low CTC-burden models. In fact, CTM were found in all models, although at a higher frequency in the MDA-MB-231 and, similarly to what happens in the clinical context, at reduced frequency compared to the whole CTC population. Beside low CTC and CTM load, systematic examination of FFPE sections from different organs confirmed also a weak metastatic potential in models with low-rate proliferating tumors, in agreement with their low CTC-burden. Assessment of metastasis formation by IHC analysis of COX IV using a species-specific antibody not cross-reacting with the murine counterpart, allowed a reliable identification of metastases even at single cell resolution. In particular, the site of metastases was different among the various models, as for instance clusters or foci consisting of few cells were found in the ovaries of MDA-MB-453 and MDA-MB-468 models only, whereas lymph-nodal involvement was exclusively observed in the MDA-MB-468 and MDA-MB-231 models. All cell lines were instead able to give rise to pulmonary metastases, although at different extent as single scattered cells or small foci

consisting in about 3-30 cells only could be observed in low CTC-burden models compared to metastatic nodules in MDA-MB-231.

In the literature, MDA-MB-231 cell were proven to induce lung metastases when injected in the tail vein of nude mice (Fraker et al., 1984), and transplantation experiments in the m.f.p. allowed ranking them among the most aggressive BC cell lines (Price et al., 1990; Zhang et al., 1991). Since the first reports, studies employing such cell line have started to proliferate and even nowadays they represent a large fraction of the literature on BC metastasis biology. Experiments using variants of the human MDA-MB-231 breast cancer cell line, capable of spontaneous metastatic spread following several rounds of *in vivo* selection (Saxena and Christofori, 2013; Munoz et al., 2006; Francia et al., 2008) were also reported. However, despite their popularity, the majority of these experimental models do not recapitulate the entire tumor progression process as in many cases metastases formation is induced by intra-venous or -arterial injection, thus providing a model for the dissection of the late phases of the metastatic cascade, but by-passing an important step of early dissemination.

Moreover, although a reasonable number of BC metastases models is now available, relevant models for some of the BC molecular subtypes are still lacking. ER-positive cell lines such as MCF7, T47D, and BT-474 are able to form tumors only in the presence of an exogenous source of estrogen, whereas, MDA-MB-468 and MDA-MB-453 take rate is not dependent on hormones. However, despite the metastatic origin of these cell lines, they have limited ability to invade and metastasize (Lacroix and Leclercq, 2004), unless subjected to selection of hormone-resistant variants or genetically modified (Clarke, 1996).

More recently, severely immunocompromised mice, such as NSG and $Rag2^{-/-}$ $\gamma c^{-/-}$ models, which exhibit T cell, B cell and NK cell immunodeficiency, were proposed to create new metastatic models. This time, MCF7 were able to give rise to metastases at lymph-node, lung, spleen and, sporadically, even at renal level when injected in the m.f.p. of NSG mice (Iorns et al., 2012). BT-474 cells were instead less metastatic in these mice, generating macro-metastases in only a few cases (axillary lymph node in 17% of mice, and spleen in 8%

of mice). The latter result is in contrast with our model where, although no macrometastases were found, small foci were detected in all animals and in all FFPE sections analyzed. In another study, use of bioimaging instruments allowed detecting multi-organ metastases in $Rag2^{-/-}$ $\gamma c^{-/-}$ mice injected at orthotopic level with MDA-MB-453 and BT-474 cell lines (Nanni et al., 2012). Therefore, a comparison of our results with information reported in literature suggests that although all our models enable the study of the entire metastatic process, spanning from PT formation and growth to metastatic colonization, not all of the BC metastatic sites observed in the clinical setting, as brain and bones, were represented. This is mainly the case with bone metastases, which are instead frequently observed in patients.

First attempts to explore hematogenous dissemination in BC experimental models were made only recently. In a technical paper published in 2008 (Eliane et al., 2008), different approaches for blood collection were tested to isolate CTCs from tumor-bearing mice, finally demonstrating that the cardiac puncture represents the most suitable approach to reach high yields without interference from contaminating normal murine epithelial cells. The authors also validated a method to enumerate CTC by applying a modified version of an *in vitro* diagnostic system for quantifying CTC in patients, obtaining numbers of CTC ranging from ~100 to 1,000 per milliliter of blood. In line with our results, the reported CTC concentration in the blood of MDA-MB-231 xenograft models was highly variable among different animals. Concerning CTC variability, despite a wide range of cells detected in this model, we have observed correlation between CTC load and tumor burden, whereas literature data from experiments with GFP-expressing MDA-MB-231 cells suggested that the PT size is not a strong indicator of CTC load (Juratli et al., 2014). Such remarkable variability in CTC load in experimental models could also be the result of fluctuations in CTC release, as suggested by a paper on a melanoma CTC-model (Juratli et al., 2014). Here the authors performed a real-time continuous monitoring of CTCs and could estimate a release of 0 to 54 CTCs every 5 minutes, also alternated to CTC-free phases.

Different results were observed in a MDA-MB-468 CTC-model (Bonnomet et al., 2012), where CTCs were detectable as early as 8 days after injection and increased 36 days later,

after which their levels remained quite constant. Differently from data obtained in the MDA-MB-231 CTC-model (Juratli et al., 2014), this peak in CTC level mirrored the increase in tumor growth at the same time point, suggesting again that the number of CTCs may indeed correlate with tumor size. In line with our results, CTCs numbers were markedly lower (ranging from 4 to 27 per blood sample) compared to the MDA-MB-231 model.

An important issue in CTC studies is represented by the timing of dissemination. We performed a time-course experiment to define a temporal window for isolation and characterization of early CTCs, where by “early CTC” we define those CTCs in a dormant state which have not yet completed the metastatic cascade. Collecting and characterizing such CTC fraction is crucial to understand the molecular steps required for early CTC survival in a quiescent state, and to unravel the molecular mechanisms behind tumor cell re-awakening and metastatic outgrowth. The MDA-MB-231 model, probably due to its pronounced aggressiveness, did not allow defining an optimal time point to isolate CTCs before they were able to colonize distant organs, since in a time-course experiment CTC release and increase during time mirrored the onset, frequency and extent of pulmonary metastases in matched FFPE sections. On the contrary, dissemination via lymphatic system was observed at the earliest time points (after 35 days from injection), suggesting that different molecular mechanisms are required to enter the lymphatic system compared to blood vessels. In agreement with this latter observation, also another study with MDA-MB-231 model reported that metastatic foci at lymph-nodes can be observed two weeks after orthotopic inoculation (Juratli et al., 2014).

Taken together, the results of our dynamic studies with the MDA-MB-231 model indicate that CTCs can be observed in the blood of animals after long period from intravenous injection as they can be actively released from pulmonary metastases. At the same time, once in the bloodstream, MDA-MB-231 cells, either intravenously injected or disseminating from the PT as CTCs, are able to rapidly reach peripheral districts and colonize the lung. This behavior might have important implications for CTC profiling studies as the CTC population isolated at animal sacrifice should contain tumor cells released by the

PT and lung metastases. Nevertheless, the possibility that the two putative CTC populations could activate or de-regulate the same pathways irrespective of the site of origin is also plausible. It is possible that studies on CTC dormancy require a different experimental model displaying specific biological characteristics, such as the one reported by Gao and colleagues (Gao et al., 2012). In that work the authors used murine BC cell lines endowed with differential metastatic potential, suggesting that time-course experiments with low CTC-burden BC cell lines might represent a solution. Indeed, also Bonnomet and colleagues found lung metastases at later time point only, despite CTC recovery was possible even few days after cell injection in a time-course experiment with the MDA-MB-468 model (Bonnomet et al., 2012). Moreover, a higher number of animals should be analyzed per experiment in order to overcome the variability in CTC load and recovery, and the support of bioimaging instruments would facilitate CTC dissemination and metastasis formation monitoring in longitudinal studies (Bonapace et al., 2012; He et al., 2007).

Further interesting data which emerged from our study is the detection of CTM in all our models. Overall, to our knowledge, the presence of CTM in experimental TN and non-TN BC models has not been reported yet, except for the LM2-MDA-MB-231 (Aceto et al., 2014) and MDA-MB-435 (Glinsky et al., 2003) xenografts, probably as a consequence of the technical approach used for CTC isolation. In our preliminary study we were instead able to monitor CTCs by size-based isolation supports and direct analysis of cytological samples which allowed detecting CTC clusters in different BC models.

After having carefully explored the limits of the experimental models with low CTC/metastases-burden, we decided to focus our molecular and biological analyses on the highly aggressive MDA-MB-231 CTC-model. Extensive molecular characterization of tumor cells in xenograft models relies on the availability of sufficiently sensitive and specific methodological approaches for their isolation, detection and molecular analysis. Since biological samples obtained from mouse xenograft models inevitably contain both human and murine cells when the method used for tumor cell isolation does not guarantee 100% purity, we designed a strategy for specific human cell detection and quantification and we

assessed that the microarray platform was sufficiently sensitive (Fina et al., 2015a) and specific to obtain reliable GEPs. The main hypothesis we have tested is that the molecular profile of CTCs is different from those of the PT and metastases. The results of the first GEP experiment, obtained in 2013, revealed for the first time (Fina et al., 2014), that experimentally obtained CTCs have a distinctive transcriptome compared to the solid lesions they give rise to or from which they originate, and that they do not retain all the molecular characteristics of the parental cell line. Importantly, this data was confirmed in an independent xenograft experiment and was furthermore corroborated after having excluded the possibility of artifacts generated from technical and experimental variability. Such analysis suggested that molecular differences observed among sample classes (PT, CTC, LN, LUNG and MDA-MB-231) are indeed mainly imputable to the tissue source, either solid tissues or blood or *in vitro* cultured parental cells, and are not simply the results of batch effects or individual variability. Differently from systemic disease the molecular profiles of transplanted cells, which gave rise to mammary nodules in two different sites, and cells disseminated via the lymphatic system or colonizing the lung were widely similar among them, thus indicating that after hematogenous or lymphatic dissemination tumor cells that establish distant lesions tend to recapitulate the characteristics of the PT. Overall, these observations demonstrate that CTCs share few properties with the source from which they originate (either the PT or pulmonary metastases, if considering the ability to re-disseminate or re-seed), while the acquisition of new molecular traits is needed during dissemination.

The achievement of such a result is the outcome of the methodological approach we have proposed, based on the combination of two CTC-enrichment approaches, a first positive selection step with immunomagnetic beads and a second size-based isolation step with dedicated filters, in order to avoid selection biases and maximize the CTC capture, followed by indirect quantification of tumor cells, either deriving from blood or solid tissues, and by gene expression profiling using species-specific molecular assays, paying attention to ensure equal inputs of human RNA in mixed murine-human samples. The outcome of preliminary quality control tests and feasibility experiments to isolate and characterize tumor

cells from blood and other tissues in our experimental models confirmed the robustness of each step in the proposed experimental workflow. Finally, large sample size (10 animals per experiments) also facilitated the selection of the best candidates for performing GEP analyses. Therefore, the usefulness and adequacy of our approach has been also demonstrated.

The extensive molecular characterization of different tumor sites and comparison analyses allowed identifying about 500 genes differentially expressed between CTCs and all solid lesions, indicating that important transcriptional reprogramming occurred during the hematogenous dissemination step or circulation in blood. Genes up- and down-regulated in CTCs are supposed to be involved not only in cell migration or intravasation, but also in resistance to *anoikis*, shear stress and attacks from the immune system. Gene Ontology analysis of genes exclusively up-regulated in CTCs identified a series of terms related to embryogenesis and development, suggesting that CTCs might be able to dedifferentiate, thus recapitulating fundamental steps of tissue migration or remodeling, and to acquire plasticity features in order to adapt to foreign environments. According to our hypothesis, up-regulated genes were also enriched in GO terms related to cell surface receptor signaling pathways and response to external stimuli and pathogens, suggesting that CTCs might modify their phenotype to react against external cues encountered during their trip. Instead, genes exclusively down-regulated in CTCs were mainly related to chromatin remodeling, packaging and methylation in particular, and negative regulation of transcription and RNA metabolism, which is consistent with the substantial transcriptional reprogramming we observed and that is supposed to orchestrate the activation of molecular pathways enabling tumor cell dissemination, survival in blood and homing to distant organs.

CTCs isolated from MDA-MB-231 xenografts were already proven to have peculiar behavior compared to the parental cell line in response to different microenvironmental conditions (Ameri et al., 2010). Ameri et al. demonstrated altered response to hypoxia, assessed as induction of ATF3, ATF4 and its target gene *ASNS*, after CTC culture under hypoxia conditions, and subsequent higher tumorigenicity compared to parental cells. More

recently the analysis of transcripts related to several metabolism pathways in CTC, PT and lung metastases from a 4T1 mouse model demonstrated an increased activation of oxidative phosphorylation pathways and enhanced mitochondrial biogenesis in CTCs compared to the other samples, while glucose metabolism was not deregulated (LeBleu et al., 2014). These data provided evidence for an opportunistic switch from an aerobic glycolysis to an oxidative metabolism in metastatic cells in order to fulfill the increased energetic demand of CTCs; it also supports the importance of characterizing CTCs to find molecular traits which distinguish them from PT and metastases. Recent data support the role of other enzymes involved in cell metabolism, as the acetyl-CoA synthetase 2, which was proven to promote acetate utilization and maintain cancer cell growth under metabolic stress in an experimental model of prostate cancer (Schug et al., 2015). Consistently, in our model gene encoding the mitochondrial fission 1 protein (*FIS1*) emerged among those significantly up-regulated in CTCs compared to solid lesions in both GEP experiments, while genes encoding the hesokinase 2 (*HK2*), the first rate-limiting enzyme of glycolysis, and the pyruvate dehydrogenase kinase (*PDK1*), involved in the regulation of pyruvate conversion to acetyl-CoA, were found to be down-regulated in CTCs, whereas acetyl-CoA synthetase 1 (*ACSS1*) was classified among genes up-regulated in CTC compared to the PT but not differentially expressed between CTC and the parental cell line. More extensive molecular analysis of CTCs isolated from mouse melanoma models allowed identifying a CTC-derived signature correlated with invasiveness and cellular motility in human melanoma after unsupervised clustering, thus providing first evidence that tumor cells committed to dissemination undergo an important transcriptional reprogramming (Luo et al., 2014). However, in this study the authors were not able to confirm data also in a second independent experiment, suggesting that the experimental variability might influence the outcome of the experiments and twist the biological interpretation of results.

Our approach can be applied to basic studies on CTC biology employing commercially available cell lines, as also primary cultures or clinical specimens (either solid biopsies or CTCs) engrafted in immunocompromised mice. One limit of the MDA-MB-231 model is

however the low representativeness of the basal B molecular subtype in the clinical setting (Neve et al., 2006). Nevertheless, seminal studies with MDA-MB-231-derived cell lines showing tropism toward specific organs unveiled the genetic determinants responsible for pulmonary metastases in BC and demonstrated their clinical relevance in predicting distant metastasis-free survival (Minn et al., 2005). Understanding metastasis initiation and progression is also critical to find genes responsible for the molecular switch from a dormant to a proliferating state, which is critical for the development of new drugs able to arrest the metastatic cascade at the earliest phases, especially if considering that CTCs can be detected even in patients with DCIS or with T_1-T_2 invasive BCs (Franken et al., 2012). Single-cell studies with BC PDX models at early and late phases of metastasis (Lawson et al., 2015) reported that tumor cells isolated from low-metastatic burden tissues prevalently expressed stem cell, EMT, pro-survival, and dormancy-associated genes, whereas metastatic cells from high-burden tissues were similar to PT cells, which were characterized by a luminal differentiation phenotype. Our model did not however allow tracing molecular changes and CTC plasticity during time as the number of cells detected at early time points in a time-course experiment were under the technical detection limit of the DASL assay platform. More importantly, once in the bloodstream, they were able to rapidly reach pulmonary capillaries and form metastatic foci.

Data emerged from the molecular characterization of experimentally derived CTCs confirmed the importance of plasticity for a tumor cell that is starting the metastatic process. One of the most debated aspects in tumor dissemination concerns the activation of an epithelial-to-mesenchymal transition process, which besides being linked to the acquisition of a stemness-like and drug-resistance phenotype enables tumor cells to migrate, invade and enter the bloodstream. In our model no shift toward a different phenotype, either pure epithelial or transient or more markedly mesenchymal, was observed in CTC compared to the PT when considering the expression of EMT-related genes, such as *SLUG*, *SNAIL*, *VIM*, *CDH1* and others, probably because MDA-MB-231 have a pronounced mesenchymal phenotype. However, similarly to the data published by Ocaña and colleagues (Ocaña et al.,

2012), CTCs were found to undergo significant down-regulation of the EMT-inducer homeobox transcription factor *PRRX1*, thus indicating that different genes and pathways can be activated, alternatively to the commonly assessed EMT-markers. Despite the fact that recently published evidence rejected the hypothesis that EMT is needed for pulmonary metastases in BC (Fischer et al., 2015), our data suggest that a reversion to an epithelial phenotype mediated by the down-regulation of *PRRX1* gene might be needed to complete the metastatic cascade, and that a subpopulation of MDA-MB-231-derived CTCs re-expressing epithelial genes may exist. The concordance between results obtained in our laboratory and those published by Ocaña (Ocaña et al., 2012), using two different experimental settings, furthermore corroborates data in favor of a metastatic model which requires the reversion from a mesenchymal to an epithelial phenotype. In our model, no modulation of the *EPCAM* gene was observed. However, it should be considered that EpCAM, although detectable at mRNA level, is expressed at low density on the surface of MDA-MB-231 cells, making it undetectable by standard antibodies and, consequently, once spiked in blood samples, MDA-MB-231 cells cannot be recovered using EpCAM-based approaches. A possible impairment of EMT, following *PRRX1* down-modulation, could be hypothesized in our CTC-models. This implies that a possible shift toward the acquisition of epithelial features could occur during dissemination. If so, we should be able to observe a re-expression of EpCAM at functional level in CTCs by *in situ* protein analysis. However, studies with MDA-MB-231 xenografts report discordant data regarding the possibility to capture CTCs using EpCAM-based approaches. The AdnaTest approach based on EpCAM/MUC1-dependent enrichment failed to detect CTCs in MDA-MB-231 xenografts in a spiking experiment, notwithstanding in another study the application of EpCAM-based CTC enrichment devices, as the MagSweeper (Ameri et al., 2010), allowed detecting and culturing CTCs derived from MDA-MB-231 xenografts. In line with this latter observation, in a preliminary test to evaluate the efficacy of our CTC enrichment approaches we could detect CTCs in blood samples from MDA-MB-231 xenografts processed with the AdnaTest EpCAM/MUC1-based approach.

CTCs may be regarded as a sort of “black-box” worth being deciphered in order to find new metastasis-associated genes for basic and pharmacological studies. The identification of candidate genes, emerging from the molecular analysis of our experimental model, worth being investigated in *in vitro* and *in vivo* assays, was driven by the hypothesis that genes up-regulated in CTCs compared to the PT are important to permit the entrance and survival in blood. Systematic analysis of gene lists, obtained according to multiple comparison between PT and CTC, taking into account also the expression level in the parental cells in order to allow preliminary evaluations *in vitro* by loss-of-function studies, led us to the definition of a panel of candidate genes significantly up-regulated in CTC compared to the PT and not significantly differentially expressed between CTC and the parental cells. Among them, *TFF3* gene, encoding for a secreted peptide of the trefoil factors family, was selected to study its role in metastasis.

We demonstrated that TFF3 is specifically involved in cell motility and invasiveness since significant impairment of migration and invasion potential was obtained after transient or stable TFF3 knock-down in the MDA-MB-231 cell line, as also in cell clones obtained following CRISPR/Cas9-mediated gene knock-out. On the contrary, MDA-MB-231 cell proliferation rate remained unvaried, thus indicating that *TFF3* can be numbered among metastasis-related genes. However, the hypothesis that TFF3 could also drive the vascular-mimicry ability, typical of MDA-MB-231 and other triple-negative cells, was not confirmed, thus suggesting that while TFF3 is important for motility and invasion, the activation of TFF3-independent pathways is required for the formation of tumor cell-lined channels. It is also plausible that cells do not have the same role and cooperate in different ways to enable tumor initiation and progression. Consistently with this view of cancer as an organized, though aberrant tissue consisting of cell populations committed to different biological functions, studies with MDA-MB-231 clones selected after *TFF3* knock-out by the CRISPR/Cas9 system revealed that cells are highly heterogeneous as the proliferation or migration potential of a single clone did not mirror the behavior of the entire cell population, irrespectively from the TFF3 gene status. Indeed, although results of *in vitro* assays obtained

in experiments with gene knock-down were again confirmed, both *TFF3* wild-type and knock-out clones showed high variability in proliferation rate, and remarkably the migration potential of wild-type clones was significantly reduced compared to the parental cells. Moreover, cell clones with unexpected abnormal behavior were observed during migration experiments.

In an attempt to assess the actual role of TFF3 in dissemination and metastasis formation, three wild-type and knock-out clones were injected in mice. Two out of 3 clones, either wild-type or knock-out, failed to initiate a tumor, and those clones able to take in the animal showed a slower growth rate compared to MDA-MB-231 parental cells. Similar results were obtained also after injecting pooled cell clones, on the hypothesis that cell cooperation could obviate to biological heterogeneity; such an attempt, however, failed, maybe because the representativeness of the entire population was too low. Animals injected with the wild-type or knock-out clones which were able to give rise to tumor nodules were sacrificed at later time points compared to the standard experimental scheme for MDA-MB-231 xenografts and, in agreement with the results obtained in *in vitro* assays, their growth rate, although similar, was slower compared to the parental cells. Despite the fact that CTCs were not detected in the blood of mice injected with the $TFF3^{KO}$ clone, both wild-type and knock-out clones were able to metastasize at lung. Moreover, contrarily to what expected on the basis of the outcome of *in vitro* preliminary experiments, the $TFF3^{KO}$ clone exhibited increased metastatic potential as pulmonary foci were larger than those originated from the $TFF3^{WT}$ cells.

At present, few data are available on the role of TFF3 in BC. *TFF3* - and also *TFF1*, but not *TFF2* - mRNA was detected in breast tumors and estrogen-responsive BC cell lines (May and Westley, 1997; Poulsom et al., 1997). In a comprehensive transcriptome and proteomic profile analysis of normal breast, noninvasive and invasive BC cell lines *TFF3* was ranked among the genes down-regulated in MDA-MB-231 compared to MCF7 cells (Nagaraja et al., 2006). In another study *TFF3* was found to be up-regulated in BT-474 and MDA-MB-361, but not SK-BR-3, HER2-positive BC cell lines compared to the HER2-negative MDA-MB-435 and

MDA-MB-468 BC cell lines, which failed to express *TFF3* in Northern blot analysis (Wilson et al., 2002), as also observed in MDA-MB-231 cells by semi-quantitative PCR (May and Westley, 2015). Putative interaction with HER2 was also confirmed in another study showing that inhibition of HER2 induces TFF3 down-regulation. Differently from data reported in literature, our GEP experiments and biochemical analyses demonstrated that TFF3 secreted peptide is detectable at mRNA level by TaqMan-based qPCR and in the conditioned medium of MDA-MB-231 using an ELISA test specific for TFF3 and which does not detect TFF1, TFF2 and other secreted molecules, suggesting that the low sensitivity or specificity of the methods used in previous studies might have caused false negatives. The expression at functional level of TFF3 was unexpected if considering its correlation with ER. In fact, in agreement with the large part of literature data, the expression of *TFF3* was observed to be induced in an estrogen-dependent and progesterone-independent manner (May and Westley, 1997) and that it is increased in tamoxifen resistant MCF7 cells (Kannan et al., 2010). Consistently, forced expression of TFF3 in ER-positive cell lines additively increased ER transcriptional activity in the presence of estradiol, promoted estrogen-independent growth and induced resistance to tamoxifen and fulvestrant (Kannan et al., 2010). In line with this evidence, we have found that *TFF3* is significantly overexpressed in MCF7-derived mammospheres, which were proven to be non-responsive to estradiol and fulvestrant (Callari et al., 2016) and to have high tumorigenicity (Ponti et al., 2005), compared to the parental cell line.

In our study, *TFF3* was found to be significantly up-regulated in MDA-MB-231-derived CTCs compared to the PT, and not significantly differentially expressed compared to the parental cell line, despite a trend toward an increased expression in CTCs. This might imply that MDA-MB-231 cell growth in the m.f.p. of mice and the establishment of a tumor exerted the function of a biological filter, which determined the selection of specific subsets of cells. The molecular analysis of disseminated cells allowed identifying genes which are responsible for MDA-MB-231 aggressive behavior and that probably could not have been discovered using different experimental designs. Therefore, also considering CTC plasticity, it is not

surprising that triple-negative cells could exploit molecular pathways generally proven to be fundamental in other BC subtypes.

Further functional *in vitro* studies revealed that forced expression of TFF3 in ER-positive BC cells increased proliferation and survival, enhanced anchorage-independent growth, promoted migration and invasion (Kannan et al., 2010, Pandey et al., 2014) and stimulated endothelial cells adhesion, transmigration through an endothelial barrier and *de novo* angiogenesis (Pandey et al., 2014; Lau et al., 2015). Similarly, we have demonstrated that MDA-MB-231-derived CTCs overexpress *TFF3* compared to the PT, and that the migration and invasion ability assessed *in vitro* after gene modulation in the parental cells is dependent on the expression of TFF3, which instead is not involved in proliferation or vascular-mimicry ability.

More studies with animal models are however needed to better decipher the role of TFF3 in BC, although it was reported in the literature that forced TFF3 expression in MCF7 cells induces the formation of metastatic foci at lung (Pandey et al., 2014). The outcome of our experiments did not allow drawing conclusion on the metastatic ability of MDA-MB-231 cells after *TFF3* knock-out, notwithstanding the hypothesis that clonal heterogeneity and/or gene compensation events could have produced biased results is plausible. Tumor heterogeneity is a well-known issue in BC, although many of its aspects still need to be clarified, and both inter-patient and intra-patient heterogeneity is observed at genomic, transcriptional and proteomic level (Polyak, 2011). In the past, seminal studies by Fidler had revealed that cell clones with different metastatic potential co-exist within a tumor cell line (Fidler, 1978; Fidler and Kripke, 1977; Kripke et al., 1978; Fidler, 2016). More recently two distinct populations were observed in MDA-MB-231 in culture, one with low surface levels of various chemokine receptors and a second with much higher levels, the latter showing increased metastatic ability in *in vivo* experiments (Norton et al., 2015). Moreover, there is growing evidence that cancer cells behave as communities and that the level of variability and cooperation among subclones can influence disease progression (Tabassum and Polyak, 2015). Recent analysis of clonal subpopulations in MDA-MB-231 and CN34 cell lines

revealed inter-cell transcript expression variability, which may enhance the cell fitness under changing environmental pressures encountered during cancer progression, including the tendency to metastasize (Nguyen et al., 2016). On the basis of these observations and given an established heterogeneity in clinical and experimental tumors, we can assume that the biological behavior of MDA-MB-231 cells is the result of a cooperation among different subsets of cells within the same population, and that the biological effects observed in *in vitro* and *in vivo* assays following artificial modulation of specific genes may be the consequence of an impairment of the cross-talk among different cell clones caused by previous *in vitro* selection procedures. Surprisingly, one of the MDA-MB-231 *TFF3*^{KO} clones selected in our studies, although validated by Sanger sequencing and ELISA test, showed extraordinary migration ability compared to the parental cell line. This might suggest that, behind clonal heterogeneity, a gene compensation effect could be hypothesized, as demonstrated in different experimental models where cells showed the ability to compensate the loss of one gene with the aberrant expression of other genes (Rossi et al., 2015). We have to consider that the CRISPR/Cas9 system can provoke alterations in many genes beyond the desired target, although its function is mediated by sgRNA designed to be specific for the gene of interest. In order to overcome this issue, *in vivo* validation experiments have recently started and are now in progress using TFF3 stable knock-down cells.

The applicability of biologically relevant data obtained from experimental models depends on the outcome of validation studies in the clinical context. In order to evaluate the clinical significance of the molecular signatures identified in our model, the expression of genes classified as CTC-specific was assessed in CTCs isolated from clinical samples and correlated with the clinical outcome. However, a plethora of methods to isolate, identify and characterize CTC exists, and the message coming from CTC analysis may influence the outcome of a clinical study according to the methodological approach. For this reason, CTC changes were firstly monitored in two independent BC case series, both including patients at different stages of disease, using the AdnaTest positive selection-based or the ScreenCell size-based isolation approaches. To our knowledge, this is the first study that compares a

positive selection-based with a size-based isolation approach in BC (Fina et al., 2015b). CTC enrichment based on standard EpCAM and MUC1 surface markers failed to detect CTCs in the vast majority of M0 BC patients (about 70%), whereas roughly 50% CTC positivity rate was obtained in M+ cases. Considering that not all CTCs express epithelial markers and that subsets of CTCs with transient epithelial-to-mesenchymal or completely mesenchymal phenotype can be generated during the dissemination phase or selected as a consequence of the adaptation to therapy, part of the case series was in parallel analyzed using another AdnaTest kit to capture CTCs, based on additional markers (ERBB2 and EGFR) beyond the classical one. The alternative CTC enrichment method showed an improved biological sensitivity as it allowed detecting CTCs in a higher number of samples compared to EpCAM/MUC1-based capture, especially in patients without clinical evidence of metastatic disease. Interestingly, CTC positivity rate was mainly driven by the detection frequency of EMT- and stemness-related genes rather than the epithelial/tumor-associated ones in M0 cases. In fact, all CTC positive samples in this clinical setting were negative for the expression of *EPCAM*, whereas the higher contribute to CTC positivity rate (more than 50%) was imputable to the expression of genes encoding ALDH1 and Akt2. Taken together, these results are consistent with the hypothesis that CTCs released in the early stage of BC progression have a marked mesenchymal/stemness-like phenotype, which might be responsible for CTC long-term dormancy and ability to resist to therapy, compared to CTCs detectable in patients with metastatic disease. Indeed, although *ALDH1A* and *AKT2* were the most represented markers in the overall case series, disseminated cells isolated from metastatic patients expressed *EPCAM* in 30% cases, suggesting that CTCs, differently from those released at earlier stages of disease, might have undergone a reversion in the EMT process at the end of their trip in order to colonize distant sites. Overall these data indicate that subsets of CTCs can be missed when using certain markers for their detection, especially in studies including patients with early-stage disease.

These observations are corroborated by the results obtained in further case series, including M0 and M+ patients, analyzed for CTC status and CTC load by a size-based

approach. In this case CTC capture is not dependent on the expression of surface markers and allows appreciating the CTC heterogeneity. ScreenCell called as positive the vast majority of cases (roughly 80%), irrespectively of the stage of disease. This data confirms that in BC tumor cells are endowed with high disseminating potential and that hematogenous dissemination is a frequent event also at the earliest phases of tumor progression, as already observed in previous experimental and clinical studies. Moreover, this approach allowed appreciating the typical CTC pleomorphism and the tendency to establish interactions, either homotypic or heterotypic, with other cells. Actually, we observed both single CTCs and CTC clusters, although at different frequency according to the clinical stage. CTC clusters, known to have increased metastatic potential compared to single CTCs according to recent reports, were surprisingly observed at higher frequency in non-metastatic patients. Moreover, also CTM total load and CTM proportion in relation to the entire CTC population were considerably high compared to M+ patients, suggesting that tumor cells preferentially disseminated as aggregates rather than single cells in this early stage. However, it should be considered that the nature of these clusters should be confirmed in light of the genetic instability as a fundamental hallmark of cancer cells as a recent single cell analysis of clusters in non metastatic colorectal cancer patients demonstrated that a discrete population of these cells is not cancerous but consists of tumor-derived circulating endothelial cells (Cima et al., 2016).

After exploring the CTC detection rate with both methods, data were analyzed in relation to the clinico-pathological features of the case series. The solely CTC status by AdnaTest, assigned on the basis of the expression of a limited panel of molecular markers, was not able to predict progression-free survival. Moreover, CTC status did not provide useful information on response to neoadjuvant therapy, probably due to the low pCR frequency in this clinical setting and the low number of cases enrolled in our study. However, CTCs seems to be associated to the HER2 status as a trend toward a significantly higher positivity was obtained in patients with HER2-positive M0 BC. Similarly to the results obtained with the AdnaTest positive selection-based approach, also CTC and CTM counts

were not associated to standard BC prognostic factors, except for the tumor size and the HER2 status. Indeed, CTM were more frequently detected in patients with T_1/T_2 or with HER2-positive primary BCs. This data corroborates the hypothesis that dissemination is an event that requires homotypic - and possibly heterotypic - interactions among cells when the tumor is at early stage, since stronger tendency to circulate as clusters rather than single cells was observed. It is also plausible that CTCs observed in metastatic patients derive from the metastatic site and that dissemination from the PT requires the activation of different pathways or the interaction with cells exerting a supportive role compared to cells that have already completed their trip and are newly released from metastatic sites. Indeed, the vast majority of patients did not present with metastatic disease at diagnosis and their tumor had been removed by chirurgical resection many years before progression. Neither the CTC nor CTM load was informative on response to therapy assessed by instrumental analysis in M+ patients, while significant increase in CTC count was observed in M0 patients categorized as complete or partial responders according to pathological assessment.

Overall, we can conclude that the CTC-message might change according to the methodological approach used, as different subsets of cells can be missed when using standard markers. Moreover, our data revealed that CTC release in early stage BC is more frequent than expected. Studies with the CellSearch system in patients with early BC allowed obtaining about 20% only of CTC positivity at baseline (Rack et al., 2014; Janni et al., 2016), compared to ~30% by AdnaTest and ~80% by ScreenCell in our case series.

AdnaTest failed to identify patients at high risk of relapse or progression when CTCs, isolated by the EpCAM/MUC1-based approach, were evaluated at baseline (cases processed with the improved kit were too low to perform correlation analysis). The same result was obtained in the DETECT trial for metastatic BC patients, designed to compare the prognostic impact of AdnaTest BreastCancer kit and the CellSearch system (Müller et al., 2012). However, longitudinal analysis of CTC by AdnaTest was proven to have superior prognostic value compared to CTC analysis performed at baseline only, as dynamic CTC changes assessed at early time points from the start of therapy allowed identifying patients

with significantly shorter progression-free survival both in colorectal (Musella et al., 2015) and bladder cancer (Fina et al., 2016) case series. In this thesis work, the low number of cases and missing blood samples during administration of therapy did not allow drawing conclusions on the clinical usefulness of AdnaTest for CTC longitudinal monitoring in metastatic BC.

The biological and clinical significance of CTC clusters in early BC remains to be clarified. Permanent presence of CTC clusters at different time points during therapy was associated to poor outcome in M+ BC patients (Aceto et al., 2014) and provides additional prognostic information beyond CTC enumeration alone (Wang et al., 2016). However, no data have been reported on the clinical relevance of CTM in M0 BCs, yet.

On the basis of the global outcome of our clinical studies, we hypothesized that the molecular characterization of CTCs isolated by AdnaTest using additional genes and *in situ* analyses at protein level to evaluate the composition of CTC clusters are needed and might help unveiling the real CTC message. To this aim, we assessed the clinical significance of a panel of CTC/metastasis-specific genes, including 12 genes exclusively up-regulated in CTCs, 4 genes selected for functional validation studies as up-regulated in CTCs compared to the PT and not significantly differentially expressed compared to the parental cell line (including TFF3), and 5 genes already reported in literature as related to metastasis and/or stemness: *OPN* (Minn et al., 2005), *SPARC* (Minn et al., 2005), *PTPN11* (Aceto et al., 2012), *SOX2* (Leis et al., 2012; Lee et al., 2016; Basu-Roy et al., 2015), and *YAP1* (Cordenonsi et al., 2011). From a technical point of view, we obtained encouraging results as our gene panel called positive a twofold higher number of cases compared to AdnaTest and detection rates were at least 15% for the majority of genes. Moreover, low concordance among genes was obtained when comparing samples processed by EMT1 with matched samples processed by EMT2 AdnaTest kits, thus confirming that different CTC subsets exist in the same patient. Interestingly, *TFF1* and *TFF3* genes were more frequently expressed in CTCs isolated from patients with metastatic disease, suggesting that they might be responsible for tumor

aggressiveness or confer metastasis-initiating properties. Instead, *HDAC10* and *PTPN11* were found among genes more frequently detected in CTCs isolated from M0 patients.

In addition to its classical role in regulating transcription, *HDAC10* aberrant expression was proven to suppress metastases (Song et al., 2013), promote autophagy-mediated cell survival and confer drug resistance (Oehme et al., 2013). A number of genes that prevent tumor growth at secondary sites, including those encoding KISS1, MKK4, MKK6, BHLHLB3/Sharp-1, and Nm23-H1, among others, were identified in dormant disseminated tumor cells (Sosa et al., 2011). On the contrary, tyrosine phosphatase SHP2 (*PTPN11*) was proven to promote progression and maintain tumor-initiating properties in breast cancer (Aceto et al., 2012). Given that CTCs can be present in the blood of patients who underwent mastectomy many years after primary tumor treatment (Meng et al., 2004b), the expression of *HDAC10* and *PTPN11* genes in CTCs released from clinically non metastatic tumor supports the view of dormant disseminated cells as quiescent low-proliferating cells, with stem cell-like phenotype, and displaying enhanced survival and drug resistance properties (Sosa et al., 2014).

When analyzing detection rates as a function of the molecular subtype, no significant trend toward the expression of particular sets of genes was observed, except for *GIGYF1* which was mainly detected in CTCs isolated from TN and HER2-positive BC. This result suggests that CTC features might not be influenced by the molecular subtype, and that their behavior is mainly driven by the activation and down-regulation of specific pathways involved in the dissemination process rather than subtype-specific genes, as also indicated by the discordant ESR1 or HER2 status between CTCs and PT (Aktas et al., 2011; Pestrin et al., 2009), already reported in many studies and confirmed also in our case series.

Interestingly, notwithstanding the limited and heterogeneous case series, correlation analyses with follow-up data revealed that *TFF3*, *TFF1*, *YAP1*, *SOX2* and, to a lesser extent, *ELF3* genes have prognostic role when their expression is assessed at CTC level, as patients with CTCs showing positivity for one of these genes had shorter progression-free survival compared to patients rated as negative.

According to literature data, *TFF3* detection in lymph-nodes allowed upstaging of sentinel lymph-nodes containing metastastic disease of 10% compared to the routine histological analysis (Weigelt et al., 2004). *TFF3* expression is also strongly correlated with breast cancers with metastases to bone (Smid et al., 2006) and was found among genes belonging to a specific genomic region displaying copy number gain at CTC level in patients with metastatic BC (Kanwar et al., 2015). However, analysis of a publicly available GEP datasets including node negative and untreated patients with primary BC showed that assessing *TFF3* expression level in the PT does not allow predicting disease outcome, thus indicating that the molecular characteristics of CTCs have distinct prognostic value compared to those of the PT.

The identification of BC biomarkers by high-throughput molecular analyses has rapidly increased in the latest year, but few studies were able to demonstrate the real clinical validity or the clinical utility of such biomarkers, probably because their biological significance is not usually investigated and top-down rather than bottom-up approaches are generally preferred. CTC-based clinical studies also lack demonstration of the clinical utility of CTC status assessment, as probably more attention should be paid to the molecular profile in addition to the CTC burden alone in order to exploit the 'liquid biopsy' opportunity fully.

To our knowledge, we have demonstrated for the first time that CTC-specific genes identified in experimental models and related to metastasis can represent clinically relevant biomarkers when their expression is monitored in the systemic disease rather than at PT level, thus confirming the superior prognostic value of biologically meaningful blood-borne biomarkers.

6. CONCLUSIONS

In the last decades, the view of cancer as systemic disease has changed the way of approaching to basic and clinical studies. Differently from cell-free nucleic acids and other circulating biomarkers, CTCs can be viewed as biological entities endowed with metastatic potential and holding a plethora of information, which are worth being investigated in order to broaden our knowledge on metastasis biology and to find new and clinically relevant prognostic biomarkers. This thesis demonstrates that extensive and comparative characterization of experimentally-derived CTCs and solid lesions allows identifying genes associated to the hematogenous phase of dissemination and defining molecular signatures to discriminate breast cancer patients at high risk of relapse or progression from those with better prognosis by real-time monitoring of such genes in the CTC fraction. Overall, the results of this work open up new opportunities for basic and clinical cancer research studies.

The development of a method to obtain reliable detection, quantification, visualization and gene expression profiles of CTCs and tumor cells in solid lesions in xenograft models was instrumental to the purpose of this thesis. We provided a useful tool since the species-specificity of molecular assays can be largely applied to studies with different model systems derived from human specimens, either cell lines or primary cultures or intact pieces of tumors or CTCs themselves, injected in laboratory animals. Indeed, the results of our approach confirmed that the biological model we have characterized is reproducible as it provides largely overlapping biological information in independent experiments. An ambitious perspective might be the use of species-specific assays to decipher the role of the tumor microenvironment in xenograft models when the stromal component of a tumor specimen has been progressively substituted by the murine counterpart, phenomenon which might imply a cross-species activity in support of tumor growth and dissemination.

From a biological point of view, we demonstrated that the CTC-specific gene profile is relevant to the CTC behavior if considering that data revealed an important transcriptional reprogramming during the dissemination phase. A hypothesis-guided approach led us identifying *TFF3* as candidate gene for functional studies, and the results of our *in vitro* studies demonstrated that this secreted peptide is important for cell migration and invasion

not only in the luminal subtype but also in the TN MDA-MB-231 BC cell line. However, technical and biological constraints due to the CRISPR/Cas9 system limited the possibility to draw conclusion on the role of TFF3 in dissemination. As also anticipated in few recent reports, cell lines are heterogeneous and our work also highlighted the effect of clonal selection in functional studies. A valid alternative beside the use of lentivirus-delivered short hairpin RNAs might be the delivering of an active Cas9 protein, instead of a plasmid-based expression, which reduces the persistence of the endonuclease in the cells, thus limiting potential off-target effects, and the analysis of the whole population instead of single cell-derived clones. If the role of TFF3 in metastasis will be validated *in vivo*, further experiments should be carried out to decipher the molecular network regulated by TFF3. To this aim, the possibility that TFF3 may act not only by autocrine or paracrine signaling in the blood microenvironment but also within an intracellular circuit, *i.e.* independently from its secretion, should be also considered if we assume that it may exert a biological role while cells are transiting in blood circulation. Moreover, as CTCs are supposed to be heterogeneous if considering the marked pleomorphism in the MDA-MB-231 CTC-model, the hypothesis of an association between TFF3 expression and the presence of a subset of CTCs with an epithelial phenotype, *e.g.* expressing EpCAM at protein level, should be taken into account and would deserve further investigation. Single cell studies and *in situ* analyses will represent a valid choice to explore tumor heterogeneity and to understand the composition and the functional role of tumor subpopulations at the different stages of disease progression, both in experimental models and, whenever possible, in patients.

Our work also revealed that CTC clusters represent a common feature in the dissemination phase of breast tumors, both in the experimental and the clinical context, irrespective of the metastatic potential, since they can be found in low metastatic-burden models and in M0 patients. However, considering that CTM may contain cells of different origin, *e.g.* tumor and stromal cells, an approach to decipher the identity of cells at single cell resolution by the analysis of genomic instability (*i.e.*, point mutations, copy number

alterations, or karyotyping) should be designed as it is fundamental to define the biological nature of such clusters of circulating cells unambiguously.

The clinical significance of experimentally-derived CTC-specific genes in our CTC case series shifts the attention to the importance of the panel of markers chosen for CTC detection and characterization when using classical techniques rather than 'omics' approaches, and tell us that depicting the molecular portrait of CTCs rather than counting them or assessing the CTC status only using standard markers can help refining the design of translational studies. Basically, the scenario offered by the study of CTCs is useful either when the investigated biomarkers show or not different expression pattern compared to the traditional biopsy, as gene signatures with prognostic value identified by the molecular characterization of the primary tumor can be monitored in real-time and repeatedly during patients treatment by a minimally-invasive approach in the systemic disease. In our work we also demonstrated that experimentally-derived and biologically relevant CTC-specific genes do not provide the same information when studied at the PT level, as interrogation of publicly available gene expression databases confirmed the prognostic role of one of our genes only. The message of these results is that the 'liquid biopsy' is not only an opportunity to retrieve information on the tumor biology and adaptation by a blood-test, but also to have access to a new set of information that only well-designed CTC studies could reveal. On the basis of these consideration, it is desirable that the outcome of this study will provide the rationale for the application of such an approach, based on the comparison between CTC and solid lesions profile, also to the clinical setting, with the possibility to understand if the identified genes exert a biological role in tumor progression, adaptation and metastasis and if they represent actionable targets.

From a general point of view, this data will help generating new hypotheses on the mechanisms that regulate hematogenous dissemination, and the role of further genes will be investigated in future functional studies with animal models and also CTC-cultures derived from xenografts, in order to have an illimited biological source to explore signaling pathways deregulated in CTCs and to test the efficacy of drugs in preventing transendothelial migration

in vitro and dissemination *in vivo*. The prognostic role of a series of CTC-specific genes and their modulation during patients' treatment should be validated in confirmation studies with other larger CTC and healthy donors' case series, and also extended to the analysis of CTC-subsets or single cells, with the final aim to provide a snapshot of tumor evolution and to identify new clinically useful CTC-signatures, as also therapeutic strategies based on the evolving nature of the disease.

REFERENCES

Aasmundstad TA and Haugen OA. DNA ploidy in intraductal breast carcinomas. Eur J Cancer. **1990**;26(9):956-9. PMID: 2177615

Aceto N, Sausgruber N, Brinkhaus H, Gaidatzis D, Martiny-Baron G, Mazzarol G, Confalonieri S, Quarto M, Hu G, Balwierz PJ, Pachkov M, Elledge SJ, van Nimwegen E, Stadler MB, Bentires-Alj M. Tyrosine phosphatase SHP2 promotes breast cancer progression and maintains tumor-initiating cells via activation of key transcription factors and a positive feedback signaling loop. Nat Med. **2012** Mar 4;18(4):529-37. DOI: 10.1038/nm.2645

Aceto N, Bardia A, Miyamoto DT, Donaldson MC, Wittner BS, Spencer JA, Yu M, Pely A, Engstrom A, Zhu H, Brannigan BW, Kapur R, Stott SL, Shioda T, Ramaswamy S, Ting DT, Lin CP, Toner M, Haber DA, Maheswaran S. Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. Cell. **2014** Aug 28;158(5):1110-22. DOI: 10.1016/j.cell.2014.07.013

Adams DL, Martin SS, Alpaugh RK, Charpentier M, Tsai S, Bergan RC, Ogden IM, Catalona W, Chumsri S, Tang CM, Cristofanilli M. Circulating giant macrophages as a potential biomarker of solid tumors. Proc Natl Acad Sci U S A. **2014** Mar 4;111(9):3514-9. DOI: 10.1073/pnas.1320198111

Adams S, Gray RJ, Demaria S, Goldstein L, Perez EA, Shulman LN, Martino S, Wang M, Jones VE, Saphner TJ, Wolff AC, Wood WC, Davidson NE, Sledge GW, Sparano JA, Badve SS. Prognostic value of tumor-infiltrating lymphocytes in triple-negative breast cancers from two phase III randomized adjuvant breast cancer trials: ECOG 2197 and ECOG 1199. J Clin Oncol. **2014** Sep 20;32(27):2959-66. DOI: 10.1200/JCO.2013.55.0491

Adnane J, Gaudray P, Dionne CA, Crumley G, Jaye M, Schlessinger J, Jeanteur P, Birnbaum D, Theillet C. BEK and FLG, two receptors to members of the FGF family, are amplified in subsets of human breast cancers. Oncogene. **1991** Apr;6(4):659-63. PMID: 1851551

Ahmed AR, Griffiths AB, Tilby MT, Westley BR, May FE. TFF3 is a normal breast epithelial protein and is associated with differentiated phenotype in early breast cancer but predisposes to invasion and metastasis in advanced disease. Am J Pathol. **2012** Mar;180(3):904-16. DOI: 10.1016/j.ajpath.2011.11.022

Aktas B, Tewes M, Fehm T, Hauch S, Kimmig R, Kasimir-Bauer S. Stem cell and epithelial-mesenchymal transition markers are frequently overexpressed in circulating tumor cells of metastatic breast cancer patients. Breast Cancer Res. **2009**;11(4):R46. DOI: 10.1186/bcr2333

Aktas B, Müller V, Tewes M, Zeitz J, Kasimir-Bauer S, Loehberg CR, Rack B, Schneeweiss A, Fehm T. Comparison of estrogen and progesterone receptor status of circulating tumor cells and the primary tumor in metastatic breast cancer patients. Gynecol Oncol. **2011** Aug;122(2):356-60. DOI: 10.1016/j.ygyno.2011.04.039

Albain KS, Barlow WE, Ravdin PM, Farrar WB, Burton GV, Ketchel SJ, Cobau CD, Levine EG, Ingle JN, Pritchard KI, Lichter AS, Schneider DJ, Abeloff MD, Henderson IC, Muss HB, Green SJ, Lew D, Livingston RB, Martino S, Osborne CK; Breast Cancer Intergroup of North America. Adjuvant chemotherapy and timing of tamoxifen in postmenopausal patients with endocrine-responsive, node-positive breast cancer: a phase 3, open-label, randomised controlled trial. Lancet. **2009** Dec 19;374(9707):2055-63. DOI: 10.1016/S0140-6736(09)61523-3

Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A. **2003** Apr 1;100(7):3983-8. DOI: 10.1073/pnas.0530291100

Ali HR, Provenzano E, Dawson SJ, Blows FM, Liu B, Shah M, Earl HM, Poole CJ, Hiller L, Dunn JA, Bowden SJ, Twelves C, Bartlett JM, Mahmoud SM, Rakha E, Ellis IO, Liu S, Gao D, Nielsen TO, Pharoah PD, Caldas C. Association between CD8+ T-cell infiltration and breast cancer survival in 12,439 patients. Ann Oncol. **2014** Aug;25(8):1536-43. DOI: 10.1093/annonc/mdu191

Alix-Panabières C and Pantel K. Challenges in circulating tumour cell research. Nat Rev Cancer. **2014** Sep;14(9):623-31. DOI: 10.1038/nrc3820

Alix-Panabières C, Bartkowiak K, Pantel K. Functional studies on circulating and disseminated tumor cells in carcinoma patients. Mol Oncol. **2016** Mar;10(3):443-9. DOI: 10.1016/j.molonc.2016.01.004

Allan AL, Keeney M. Circulating tumor cell analysis: technical and statistical considerations for application to the clinic. J Oncol. **2010**; 2010:426218. DOI: 10.1155/2010/426218

Alvarez RH, Valero V, Hortobagyi GN. Emerging targeted therapies for breast cancer. J Clin Oncol. **2010** Jul 10;28(20):3366-79. DOI: 10.1200/JCO.2009.25.4011

Ameri K, Luong R, Zhang H, Powell AA, Montgomery KD, Espinosa I, Bouley DM, Harris AL, Jeffrey SS. Circulating tumour cells demonstrate an altered response to hypoxia and an aggressive phenotype. Br J Cancer. **2010** Feb 2;102(3):561-9. DOI: 10.1038/sj.bjc.6605491

Anampa J, Makower D2, Sparano JA3. Progress in adjuvant chemotherapy for breast cancer: an overview. BMC Med. **2015** Aug 17;13:195. DOI: 10.1186/s12916-015-0439-8

Angeloni V, Tiberio P, Appierto V, Daidone MG. Implications of stemness-related signaling pathways in breast cancer response to therapy. Semin Cancer Biol. **2015** Apr;31:43-51. DOI: 10.1016/j.semcancer.2014.08.004

Ao A, Morrison BJ, Wang H, López JA, Reynolds BA, Lu J. Response of estrogen receptor-positive breast cancer tumorspheres to antiestrogen treatments. PLoS One. **2011** Apr 14;6(4):e18810. DOI: 10.1371/journal.pone.0018810

Apostolou P and Fostira F. Hereditary breast cancer: the era of new susceptibility genes. Biomed Res Int. **2013**;2013:747318. DOI: 10.1155/2013/747318

Artandi SE and DePinho RA. A critical role for telomeres in suppressing and facilitating carcinogenesis. Curr Opin Genet Dev. **2000** Feb;10(1):39-46. DOI: 10.1016/S0959-437X(99)00047-7

Arun G, Diermeier S, Akerman M, Chang KC, Wilkinson JE, Hearn S, Kim Y, MacLeod AR, Krainer AR, Norton L, Brogi E, Egeblad M, Spector DL. Differentiation of mammary tumors and reduction in metastasis upon Malat1 lncRNA loss. Genes Dev. **2016** Jan 1;30(1):34-51. DOI: 10.1101/gad.270959.115

Ashworth TR. A case of cancer in which cells similar to those in the tumours were seen in the blood after death". **1869**. Australian Medical Journal. 14: 146–7

Aslakson CJ and Miller FR. Selective events in the metastatic process defined by analysis of the sequential dissemination of subpopulations of a mouse mammary tumor. Cancer Res. **1992** Mar 15;52(6):1399-405. PMID: 1540948

Asselin-Labat ML, Sutherland KD, Barker H, Thomas R, Shackleton M, Forrest NC, Hartley L, Robb L, Grosveld FG, van der Wees J, Lindeman GJ, Visvader JE. Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation. Nat Cell Biol. **2007** Feb;9(2):201-9. DOI: 10.1038/ncb1530

Atchley DP, Albarracin CT, Lopez A, Valero V, Amos CI, Gonzalez-Angulo AM, Hortobagyi GN, Arun BK. Clinical and pathologic characteristics of patients with BRCA-positive and BRCA-negative breast cancer. J Clin Oncol. **2008** Sep 10;26(26):4282-8. DOI: 10.1200/JCO.2008.16.6231

Au SH, Storey BD, Moore JC, Tang Q, Chen YL, Javaid S, Sarioglu AF, Sullivan R, Madden MW, O'Keefe R, Haber DA, Maheswaran S, Langenau DM, Stott SL, Toner M. Clusters of circulating tumor cells traverse capillary-sized vessels. Proc Natl Acad Sci U S A. **2016** May 3;113(18):4947-52. DOI: 10.1073/pnas

Baccelli I, Schneeweiss A, Riethdorf S, Stenzinger A, Schillert A, Vogel V, Klein C, Saini M, Bäuerle T, Wallwiener M, Holland-Letz T, Höfner T, Sprick M, Scharpff M, Marmé F, Sinn HP, Pantel K, Weichert W, Trumpp A. Identification of a population of blood circulating tumor cells from breast cancer patients that initiates metastasis in a xenograft assay. Nat Biotechnol. **2013** Jun;31(6):539-44. DOI: 10.1038/nbt.2576

Bargmann CI, Hung MC, Weinberg RA. Multiple independent activations of the neu oncogene by a point mutation altering the transmembrane domain of p185. Cell. **1986** Jun 6;45(5):649-57. PMID: 2871941

Baselga J and Swain SM. Novel anticancer targets: revisiting ERBB2 and discovering ERBB3. Nat Rev Cancer. **2009** Jul;9(7):463-75. DOI: 10.1038/nrc2656

Basu-Roy U, Bayin NS, Rattanakorn K, Han E, Placantonakis DG, Mansukhani A, Basilico C. Sox2 antagonizes the Hippo pathway to maintain stemness in cancer cells. Nat Commun. **2015** Apr 2;6:6411. DOI: 10.1038/ncomms7411

Battaglia F, Polizzi G, Scambia G, Rossi S, Panici PB, Iacobelli S, Crucitti F, Mancuso S. Receptors for epidermal growth factor and steroid hormones in human breast cancer. Oncology. **1988**;45(6):424-7. PMID: 3186151

Beatson GT. On the treatment of inoperable cases of carcinoma of the mamma: suggestions for a new method of treatment with illustrative cases. Lancet. **1896**;ii:104–107

Belgiovine C, D'Incalci M, Allavena P, Frapolli R. Tumor-associated macrophages and anti-tumor therapies: complex links. Cell Mol Life Sci. **2016** Jul;73(13):2411-24. DOI: 10.1007/s00018-016-2166-5

Bidard FC, Kirova YM, Vincent-Salomon A, Alran S, de Rycke Y, Sigal-Zafrani B, Sastre-Garau X, Mignot L, Fourquet A, Pierga JY. Disseminated tumor cells and the risk of locoregional recurrence in nonmetastatic breast cancer. Ann Oncol. **2009** Nov;20(11):1836-41. DOI: 10.1093/annonc/mdp200

Bidard FC, Pierga JY, Soria JC, Thiery JP. Translating metastasis-related biomarkers to the clinic--progress and pitfalls. Nat Rev Clin Oncol. **2013** Mar;10(3):169-79. DOI: 10.1038/nrclinonc.2013.4

Bidard FC, Peeters DJ, Fehm T, Nolé F, Gisbert-Criado R, Mavroudis D, Grisanti S, Generali D, Garcia-Saenz JA, Stebbing J, Caldas C, Gazzaniga P, Manso L, Zamarchi R, de Lascoiti AF, De Mattos-Arruda L, Ignatiadis M, Lebofsky R, van Laere SJ, Meier-Stiegen F, Sandri MT, Vidal-Martinez J, Politaki E, Consoli F, Bottini A, Diaz-Rubio E, Krell J, Dawson SJ, Raimondi C, Rutten A, Janni W, Munzone E, Carañana V, Agelaki S, Almici C, Dirix L, Solomayer EF, Zorzino L, Johannes H, Reis-Filho JS, Pantel K, Pierga JY, Michiels S. Clinical validity of circulating tumour cells in patients with metastatic breast cancer: a pooled analysis of individual patient data. Lancet Oncol. **2014** Apr;15(4):406-14. DOI: 10.1016/S1470-2045(14)70069-5

Bidard FC, Proudhon C, Pierga JY. Circulating tumor cells in breast cancer. Mol Oncol. **2016** Mar;10(3):418-30. DOI: 10.1016/j.molonc.2016.01.001

Bièche I and Lidereau R. Genetic alterations in breast cancer. Genes Chromosomes Cancer. **1995** Dec;14(4):227-51. PMID: 8605112

Birkbak NJ, Eklund AC, Li Q, McClelland SE, Endesfelder D, Tan P, Tan IB, Richardson AL, Szallasi Z, Swanton C. Paradoxical relationship between chromosomal instability and survival outcome in cancer. Cancer Res. **2011** May 15;71(10):3447-52. DOI: 10.1158/0008-5472.CAN-10-3667

Björnström L and Sjöberg M. Mechanisms of estrogen receptor signaling: convergence of genomic and nongenomic actions on target genes. Mol Endocrinol. **2005** Apr;19(4):833-42. DOI: 10.1210/me.2004-0486

Bombonati A and Sgroi DC. The molecular pathology of breast cancer progression. J Pathol. **2011** Jan;223(2):307-17. DOI: 10.1002/path.2808

Bonadonna G, Brusamolino E, Valagussa P, Rossi A, Brugnatelli L, Brambilla C, De Lena M, Tancini G, Bajetta E, Musumeci R, Veronesi U. Combination chemotherapy as an adjuvant

treatment in operable breast cancer. N Engl J Med. **1976** Feb 19;294(8):405-10. DOI: 10.1056/NEJM197602192940801

Bonapace L, Wyckoff J, Oertner T, Van Rheenen J, Junt T, Bentires-Alj M. If you don't look, you won't see: intravital multiphoton imaging of primary and metastatic breast cancer. J Mammary Gland Biol Neoplasia. **2012** Jun;17(2):125-9. DOI: 10.1007/s10911-012-9250-8

Bonnomet A, Syne L, Brysse A, Feyereisen E, Thompson EW, Noël A, Foidart JM, Birembaut P, Polette M, Gilles C. A dynamic in vivo model of epithelial-to-mesenchymal transitions in circulating tumor cells and metastases of breast cancer. Oncogene. **2012** Aug 16;31(33):3741-53. DOI: 10.1038/onc.2011.540

Bos PD, Zhang XH, Nadal C, Shu W, Gomis RR, Nguyen DX, Minn AJ, van de Vijver MJ, Gerald WL, Foekens JA, Massagué J. Genes that mediate breast cancer metastasis to the brain. Nature. **2009** Jun 18;459(7249):1005-9. DOI: 10.1038/nature08021

Bos PD, Nguyen DX, Massagué J. Modeling metastasis in the mouse. Curr Opin Pharmacol. **2010** Oct;10(5):571-7. DOI: 10.1016/j.coph.2010.06.003

Bose R, Kavuri SM, Searleman AC, Shen W, Shen D, Koboldt DC, Monsey J, Goel N, Aronson AB, Li S, Ma CX, Ding L, Mardis ER, Ellis MJ. Activating HER2 mutations in HER2 gene amplification negative breast cancer. Cancer Discov. **2013** Feb;3(2):224-37. DOI: 10.1158/2159-8290.CD-12-0349

Bosma AJ, Weigelt B, Lambrechts AC, Verhagen OJ, Pruntel R, Hart AA, Rodenhuis S, van 't Veer LJ. Detection of circulating breast tumor cells by differential expression of marker genes. Clin Cancer Res. **2002** Jun;8(6):1871-7. PMID: 12060630

Bouchard L, Lamarre L, Tremblay PJ, Jolicoeur P. Stochastic appearance of mammary tumors in transgenic mice carrying the MMTV/c-neu oncogene. Cell. **1989** Jun 16;57(6):931-6. PMID: 2567634

Boyd S. On oophorectomy in cancer of the breast. J Br Med. **1900**;2:1161

Boyden S. The chemotactic effect of mixtures of antibody and antigen on polymorphonuclear leucocytes. J Exp Med. **1962** Mar 1;115:453-66. PMCID: PMC2137509

Brabletz T, Jung A, Reu S, Porzner M, Hlubek F, Kunz-Schughart LA, Knuechel R, Kirchner T. Variable beta-catenin expression in colorectal cancers indicates tumor progression driven by the tumor environment. Proc Natl Acad Sci U S A. **2001** Aug 28;98(18):10356-61. DOI: 10.1073/pnas.171610498

Brabletz T. To differentiate or not--routes towards metastasis. Nat Rev Cancer. **2012** May 11;12(6):425-36. DOI: 10.1038/nrc3265

Brandt B, Junker R, Griwatz C, Heidl S, Brinkmann O, Semjonow A, Assmann G, Zänker KS. Isolation of prostate-derived single cells and cell clusters from human peripheral blood. Cancer Res. **1996** Oct 15;56(20):4556-61. PMID: 8840959

Brooks MD, Burness ML, Wicha MS. Therapeutic Implications of Cellular Heterogeneity and Plasticity in Breast Cancer. Cell Stem Cell. **2015** Sep 3;17(3):260-71. DOI: 10.1016/j.stem.2015.08.014

Buache E, Etique N, Alpy F, Stoll I, Muckensturm M, Reina-San-Martin B, Chenard MP, Tomasetto C, Rio MC. Deficiency in trefoil factor 1 (TFF1) increases tumorigenicity of human breast cancer cells and mammary tumor development in TFF1-knockout mice. Oncogene. **2011** Jul 21;30(29):3261-73. DOI: 10.1038/onc.2011.41

Budd GT, Cristofanilli M, Ellis MJ, Stopeck A, Borden E, Miller MC, Matera J, Repollet M, Doyle GV, Terstappen LW, Hayes DF. Circulating tumor cells versus imaging--predicting overall survival in metastatic breast cancer. Clin Cancer Res. **2006** Nov 1;12(21):6403-9.

Burstein HJ. The distinctive nature of HER2-positive breast cancers. N Engl J Med. **2005** Oct 20;353(16):1652-4. DOI: 10.1056/NEJMp058197

Calcagno AM, Salcido CD, Gillet JP, Wu CP, Fostel JM, Mumau MD, Gottesman MM, Varticovski L, Ambudkar SV. Prolonged drug selection of breast cancer cells and enrichment of cancer stem cell characteristics. J Natl Cancer Inst. **2010** Nov 3;102(21):1637-52. DOI: 10.1093/jnci/djq361

Callari M, Guffanti A, Soldà G, Merlino G, Fina E, Brini E, Moles A, Cappelletti V, Daidone MG. In-depth characterization of breast cancer tumor-promoting cell transcriptome by RNA sequencing and microarrays. Oncotarget. **2016** Jan 5;7(1):976-94. DOI: 10.18632/oncotarget.5810

Calnan DP, Westley BR, May FE, Floyd DN, Marchbank T, Playford RJ. The trefoil peptide TFF1 inhibits the growth of the human gastric adenocarcinoma cell line AGS. J Pathol. **1999** Jul;188(3):312-7. DOI: 10.1002/(SICI)1096-9896(199907)188:3<312::AID-PATH360>3.0.CO;2-P

Camerer E, Qazi AA, Duong DN, Cornelissen I, Advincula R, Coughlin SR. Platelets, protease-activated receptors, and fibrinogen in hematogenous metastasis. Blood. **2004** Jul 15;104(2):397-401. DOI: 10.1182/blood-2004-02-0434

Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature. **2012** Oct 4;490(7418):61-70. DOI: 10.1038/nature11412

Cappelletti V, Brivio M, Miodini P, Granata G, Coradini D, Di Fronzo G. Simultaneous estimation of epidermal growth factor receptors and steroid receptors in a series of 136 resectable primary breast tumors. Tumour Biol. **1988**;9(4):200-11. PMID: 3420376

Cardiff RD and Wellings SR. The comparative pathology of human and mouse mammary glands. J Mammary Gland Biol Neoplasia. **1999** Jan;4(1):105-22. PMID: 10219910

Cardoso F, Van't Veer L, Rutgers E, Loi S, Mook S, Piccart-Gebhart MJ. Clinical application of the 70 gene profile: the MINDACT trial. J Clin Oncol. **2008** Feb 10;26(5):729-35. DOI: 10.1200/JCO.2007.14.3222

Carpenter R, Gibbs N, Matthews J, Cooke T. Importance of cellular DNA content in pre-malignant breast disease and pre-invasive carcinoma of the female breast. Br J Surg. **1987** Oct;74(10):905-6. DOI: 10.1002/bjs.1800741011

Celià-Terrassa T and Kang Y. Distinctive properties of metastasis-initiating cells. Genes Dev. **2016** Apr 15;30(8):892-908. DOI: 10.1101/gad.277681.116

Céspedes MV, Casanova I, Parreño M, Mangues R. Mouse models in oncogenesis and cancer therapy. Clin Transl Oncol. **2006** May;8(5):318-29. PMID: 16760006

Chadwick MP, Westley BR, May FE. Homodimerization and hetero-oligomerization of the single-domain trefoil protein pNR-2/pS2 through cysteine 58. Biochem J. **1997** Oct 1;327 (Pt 1):117-23. DOI: 10.1042/bj3270117

Chaffer CL, Brueckmann I, Scheel C, Kaestli AJ, Wiggins PA, Rodrigues LO, Brooks M, Reinhardt F, Su Y, Polyak K, Arendt LM, Kuperwasser C, Bierie B, Weinberg RA. Normal and neoplastic nonstem cells can spontaneously convert to a stem-like state. Proc Natl Acad Sci U S A. **2011** May 10;108(19):7950-5. DOI: 10.1073/pnas.1102454108

Charafe-Jauffret E, Ginestier C, Iovino F, Tarpin C, Diebel M, Esterni B, Houvenaeghel G, Extra JM, Bertucci F, Jacquemier J, Xerri L, Dontu G, Stassi G, Xiao Y, Barsky SH, Birnbaum D, Viens P, Wicha MS. Aldehyde dehydrogenase 1-positive cancer stem cells mediate metastasis and poor clinical outcome in inflammatory breast cancer. Clin Cancer Res. **2010** Jan 1;16(1):45-55. DOI: 10.1158/1078-0432.CCR-09-1630

Chen Q, Boire A, Jin X, Valiente M, Er EE, Lopez-Soto A, Jacob LS, Patwa R, Shah H, Xu K, Cross JR, Massagué J. Carcinoma-astrocyte gap junctions promote brain metastasis by cGAMP transfer. Nature. **2016** May 18;533(7604):493-8. DOI: 10.1038/nature18268

Chen YH, Lu Y, De Plaen IG, Wang LY, Tan XD. Transcription factor NF-kappaB signals antianoikic function of trefoil factor 3 on intestinal epithelial cells. Biochem Biophys Res Commun. **2000** Aug 11;274(3):576-82. DOI: 10.1006/bbrc.2000.3176

Cheung KJ, Padmanaban V, Silvestri V, Schipper K, Cohen JD, Fairchild AN, Gorin MA, Verdone JE, Pienta KJ, Bader JS, Ewald AJ. Polyclonal breast cancer metastases arise from collective dissemination of keratin 14-expressing tumor cell clusters. Proc Natl Acad Sci U S A. **2016** Feb 16;113(7):E854-63. DOI: 10.1073/pnas.1508541113

Chin K, de Solorzano CO, Knowles D, Jones A, Chou W, Rodriguez EG, Kuo WL, Ljung BM, Chew K, Myambo K, Miranda M, Krig S, Garbe J, Stampfer M, Yaswen P, Gray JW, Lockett SJ. In situ analyses of genome instability in breast cancer. Nat Genet. **2004** Sep;36(9):984-8. DOI: 10.1038/ng1409

Chin K, DeVries S, Fridlyand J, Spellman PT, Roydasgupta R, Kuo WL, Lapuk A, Neve RM, Qian Z, Ryder T, Chen F, Feiler H, Tokuyasu T, Kingsley C, Dairkee S, Meng Z, Chew K, Pinkel D, Jain A, Ljung BM, Esserman L, Albertson DG, Waldman FM, Gray JW. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer Cell. **2006** Dec;10(6):529-41. DOI: 10.1016/j.ccr.2006.10.009

Cho EH, Wendel M, Luttgen M, Yoshioka C, Marrinucci D, Lazar D, Schram E, Nieva J, Bazhenova L, Morgan A, Ko AH, Korn WM, Kolatkar A, Bethel K, Kuhn P. Characterization of circulating tumor cell aggregates identified in patients with epithelial tumors. Phys Biol. **2012** Feb;9(1):016001. DOI: 10.1088/1478-3975/9/1/016001

Christofori G. New signals from the invasive front. Nature. **2006** May 25;441(7092):444-50. DOI: 10.1038/nature04872

Cima I, Kong SL, Sengupta D, Tan IB, Phyo WM, Lee D, Hu M, Iliescu C, Alexander I, Goh WL, Rahmani M, Suhaimi NA, Vo JH, Tai JA, Tan JH, Chua C, Ten R, Lim WJ, Chew MH, Hauser CA, van Dam RM, Lim WY, Prabhakar S, Lim B, Koh PK, Robson P, Ying JY, Hillmer AM, Tan MH. Tumor-derived circulating endothelial cell clusters in colorectal cancer. Sci Transl Med. **2016** Jun 29;8(345):345ra89. DOI: 10.1126/scitranslmed.aad7369

Clarke R. Human breast cancer cell line xenografts as models of breast cancer. The immunobiologies of recipient mice and the characteristics of several tumorigenic cell lines. Breast Cancer Res Treat. **1996**;39(1):69-86. PMID: 8738607

Clarke R, Liu MC, Bouker KB, Gu Z, Lee RY, Zhu Y, Skaar TC, Gomez B, O'Brien K, Wang Y, Hilakivi-Clarke LA. Antiestrogen resistance in breast cancer and the role of estrogen receptor signaling. Oncogene. **2003** Oct 20;22(47):7316-39. DOI: 10.1038/sj.onc.1206937

Clawson GA, Matters GL, Xin P, Imamura-Kawasawa Y, Du Z, Thiboutot DM, Helm KF, Neves RI, Abraham T. Macrophage-tumor cell fusions from peripheral blood of melanoma patients. PLoS One. **2015** Aug 12;10(8):e0134320. DOI: 10.1371/journal.pone.0134320

Coates AS, Winer EP, Goldhirsch A, Gelber RD, Gnant M, Piccart-Gebhart M, Thürlimann B, Senn HJ, et al. Panel Members. Tailoring therapies--improving the management of early breast cancer: St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2015. Ann Oncol. **2015** Aug;26(8):1533-46. DOI: 10.1093/annonc/mdv221

Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52,705 women with breast cancer and 108,411 women without breast cancer. Lancet. **1997** Oct 11;350(9084):1047-59. PMID: 10213546

Coradini D, Gariboldi M, Fedeli M, Pierotti MA, Daidone MG. Invasiveness gene signature predicts a favorable outcome also in estrogen receptor-positive primary breast cancers treated with adjuvant tamoxifen. Breast Cancer Res Treat. **2008** Sep;111(2):389-90. DOI: 10.1007/s10549-007-9777-3

Cordenonsi M, Zanconato F, Azzolin L, Forcato M, Rosato A, Frasson C, Inui M, Montagner M, Parenti AR, Poletti A, Daidone MG, Dupont S, Basso G, Bicciato S, Piccolo S. The Hippo transducer TAZ confers cancer stem cell-related traits on breast cancer cells. Cell. **2011** Nov 11;147(4):759-72. DOI: 10.1016/j.cell.2011.09.048

Courjal F and Theillet C. Comparative genomic hybridization analysis of breast tumors with predetermined profiles of DNA amplification. Cancer Res. **1997** Oct 1;57(19):4368-77. PMID: 9331100

Coussens L, Yang-Feng TL, Liao YC, Chen E, Gray A, McGrath J, Seeburg PH, Libermann TA, Schlessinger J, Francke U, et al. Tyrosine kinase receptor with extensive homology to EGF receptor shares chromosomal location with neu oncogene. Science. **1985** Dec 6;230(4730):1132-9. PMID: 2999974

Coussens LM and Werb Z. Inflammation and cancer. Nature. **2002** Dec 19-26;420(6917):860-7. DOI: 10.1038/nature01322

Cowin P and Wysolmerski J. Molecular mechanisms guiding embryonic mammary gland development. Cold Spring Harb Perspect Biol. **2010** Jun;2(6):a003251. DOI: 10.1101/cshperspect.a003251

Cristofanilli M, Budd GT, Ellis MJ, Stopeck A, Matera J, Miller MC, Reuben JM, Doyle GV, Allard WJ, Terstappen LW, Hayes DF. Circulating tumor cells, disease progression, and survival in metastatic breast cancer. N Engl J Med. **2004** Aug 19;351(8):781-91. DOI: 10.1056/NEJMoa040766

Cristofanilli M, Hayes DF, Budd GT, Ellis MJ, Stopeck A, Reuben JM, Doyle GV, Matera J, Allard WJ, Miller MC, Fritsche HA, Hortobagyi GN, Terstappen LW. Circulating tumor cells: a novel prognostic factor for newly diagnosed metastatic breast cancer. J Clin Oncol. **2005** Mar 1;23(7):1420-30. DOI: 10.1200/JCO.2005.08.140

Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Gräf S, Ha G, Haffari G, Bashashati A, Russell R, McKinney S; METABRIC Group, Langerød A, Green A, Provenzano E, Wishart G, Pinder S, Watson P, Markowetz F, Murphy L, Ellis I, Purushotham A, Børresen-Dale AL, Brenton JD, Tavaré S, Caldas C, Aparicio S. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. **2012** Apr 18;486(7403):346-52. DOI: 10.1038/nature10983

Daidone MG, Zaffaroni N, Cappelletti V. Strategies to translate preclinical information to breast cancer patient benefit. J Natl Cancer Inst Monogr. **2011**;2011(43):55-9. DOI: 10.1093/jncimonographs/lgr033

Dalerba P, Cho RW, Clarke MF. Cancer stem cells: models and concepts. Annu Rev Med. **2007**;58:267-84. DOI: 10.1146/annurev.med.58.062105.204854

Dawson SJ, Provenzano E, Caldas C. Triple negative breast cancers: clinical and prognostic implications. Eur J Cancer. **2009** Sep;45 Suppl 1:27-40. DOI: 10.1016/S0959-8049(09)70013-9

Dawson SJ, Tsui DW, Murtaza M, Biggs H, Rueda OM, Chin SF, Dunning MJ, Gale D, Forshew T, Mahler-Araujo B, Rajan S, Humphray S, Becq J, Halsall D, Wallis M, Bentley D, Caldas C, Rosenfeld N. Analysis of circulating tumor DNA to monitor metastatic breast cancer. N Engl J Med. **2013** Mar 28;368(13):1199-209. DOI: 10.1056/NEJMoa1213261

Deng G, Krishnakumar S, Powell AA, Zhang H, Mindrinos MN, Telli ML, Davis RW, Jeffrey SS. Single cell mutational analysis of PIK3CA in circulating tumor cells and metastases in breast cancer reveals heterogeneity, discordance, and mutation persistence in cultured disseminated tumor cells from bone marrow. BMC Cancer. **2014** Jun 19;14:456. DOI: 10.1186/1471-2407-14-456

Deome KB, Faulkin LJ Jr, Bern HA, Blair PB. Development of mammary tumors from hyperplastic alveolar nodules transplanted into gland-free mammary fat pads of female C3H mice. Cancer Res. **1959** Jun;19(5):515-20. PMID: 13663040

DePinho RA. The age of cancer. Nature. **2000** Nov 9;408(6809):248-54. DOI: 10.1038/35041694

DeRose YS, Wang G, Lin YC, Bernard PS, Buys SS, Ebbert MT, Factor R, Matsen C, Milash BA, Nelson E, Neumayer L, Randall RL, Stijleman IJ, Welm BE, Welm AL. Tumor grafts derived from women with breast cancer authentically reflect tumor pathology, growth, metastasis and disease outcomes. Nat Med. **2011** Oct 23;17(11):1514-20. DOI: 10.1038/nm.2454

Desitter I, Guerrouahen BS, Benali-Furet N, Wechsler J, Jänne PA, Kuang Y, Yanagita M, Wang L, Berkowitz JA, Distel RJ, Cayre YE. A new device for rapid isolation by size and characterization of rare circulating tumor cells. Anticancer Res. **2011** Feb;31(2):427-41. PMID: 21378321

Dexter DL, Kowalski HM, Blazar BA, Fligiel Z, Vogel R, Heppner GH. Heterogeneity of tumor cells from a single mouse mammary tumor. Cancer Res. **1978** Oct;38(10):3174-81. PMID: 210930

Diamanti-Kandarakis E, Bourguignon JP, Giudice LC, Hauser R, Prins GS, Soto AM, Zoeller RT, Gore AC. Endocrine-disrupting chemicals: an Endocrine Society scientific statement. Endocr Rev. **2009** Jun;30(4):293-342. DOI: 10.1210/er.2009-0002

Ding L, Ellis MJ, Li S, Larson DE, Chen K, Wallis JW, Harris CC, McLellan MD, Fulton RS, Fulton LL, Abbott RM, Hoog J, Dooling DJ, Koboldt DC, Schmidt H, Kalicki J, Zhang Q, Chen L, Lin L, Wendl MC, McMichael JF, Magrini VJ, Cook L, McGrath SD, Vickery TL, Appelbaum E, Deschryver K, Davies S, Guintoli T, Lin L, Crowder R, Tao Y, Snider JE, Smith SM, Dukes AF, Sanderson GE, Pohl CS, Delehaunty KD, Fronick CC, Pape KA, Reed JS, Robinson JS, Hodges JS, Schierding W, Dees ND, Shen D, Locke DP, Wiechert ME, Eldred JM, Peck JB, Oberkfell BJ, Lolofie JT, Du F, Hawkins AE, O'Laughlin MD, Bernard KE, Cunningham M, Elliott G, Mason MD, Thompson DM Jr, Ivanovich JL, Goodfellow PJ, Perou CM, Weinstock GM, Aft R, Watson M, Ley TJ, Wilson RK, Mardis ER. Genome remodelling in a basal-like breast cancer metastasis and xenograft. Nature. **2010** Apr 15;464(7291):999-1005. DOI: 10.1038/nature08989

Doane AS, Danso M, Lal P, Donaton M, Zhang L, Hudis C, Gerald WL. An estrogen receptor-negative breast cancer subset characterized by a hormonally regulated transcriptional program and response to androgen. Oncogene. **2006** Jun 29;25(28):3994-4008. DOI: 10.1038/sj.onc.1209415

Dontu G, Abdallah WM, Foley JM, Jackson KW, Clarke MF, Kawamura MJ, Wicha MS. In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. Genes Dev. **2003** May 15;17(10):1253-70. DOI: 10.1101/gad.1061803

Dorland WAN, Dorland's Illustrated Medical Dictionary 24th Edition. **1965**. Philadelphia, Saunders

dos Santos CO, Rebbeck C, Rozhkova E, Valentine A, Samuels A, Kadiri LR, Osten P, Harris EY, Uren PJ, Smith AD, Hannon GJ. Molecular hierarchy of mammary differentiation yields refined markers of mammary stem cells. Proc Natl Acad Sci U S A. **2013** Apr 30;110(18):7123-30. DOI: 10.1073/pnas.1303919110

Drukker CA, Bueno-de-Mesquita JM, Retèl VP, van Harten WH, van Tinteren H, Wesseling J, Roumen RM, Knauer M, van 't Veer LJ, Sonke GS, Rutgers EJ, van de Vijver MJ, Linn SC. A prospective evaluation of a breast cancer prognosis signature in the observational RASTER study. Int J Cancer. **2013** Aug 15;133(4):929-36. DOI: 10.1002/ijc.28082

Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. Bioinformatics. **2008** Jul 1;24(13):1547-8. DOI: 10.1093/bioinformatics/btn224

Duda DG, Duyverman AM, Kohno M, Snuderl M, Steller EJ, Fukumura D, Jain RK. Malignant cells facilitate lung metastasis by bringing their own soil. Proc Natl Acad Sci U S A. **2010** Dec 14;107(50):21677-82. DOI: 10.1073/pnas.1016234107

Eckhardt BL, Parker BS, van Laar RK, Restall CM, Natoli AL, Tavaria MD, Stanley KL, Sloan EK, Moseley JM, Anderson RL. Genomic analysis of a spontaneous model of breast cancer metastasis to bone reveals a role for the extracellular matrix. Mol Cancer Res. **2005** Jan;3(1):1-13. PMID: 15671244

Eliane JP, Repollet M, Luker KE, Brown M, Rae JM, Dontu G, Schott AF, Wicha M, Doyle GV, Hayes DF, Luker GD. Monitoring serial changes in circulating human breast cancer cells in murine xenograft models. Cancer Res. **2008** Jul 15;68(14):5529-32. DOI: 10.1158/0008-5472.CAN-08-0630

Ellis P. WHO Classification of tumours. Pathology and Genetics of Tumours of the Breast and Female Genital Organs. **2003**. Lyon Press: Lyon, 13–59

Elston CW, Ellis IO, Pinder SE. Pathological prognostic factors in breast cancer. Crit Rev Oncol Hematol. **1999** Aug;31(3):209-23. DOI: 10.1016/S1040-8428(99)00034-7

Emami S, Rodrigues S, Rodrigue CM, Le Floch N, Rivat C, Attoub S, Bruyneel E, Gespach C. Trefoil factor family (TFF) peptides and cancer progression. Peptides. **2004** May;25(5):885-98. DOI: 10.1016/j.peptides.2003.10.019

Ewing J. Neoplastic Diseases: A Treatise on Tumours. **1928**. Philadelphia, Saunders

Fan P, Fan S, Wang H, Mao J, Shi Y, Ibrahim MM, Ma W, Yu X, Hou Z, Wang B, Li L. Genistein decreases the breast cancer stem-like cell population through Hedgehog pathway. Stem Cell Res Ther. **2013**;4(6):146. DOI: 10.1186/scrt357

Fearon ER and Vogelstein B. A genetic model for colorectal tumorigenesis. Cell. **1990** Jun 1;61(5):759-67. PMID: 2188735

Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer. **2010** Dec 15;127(12):2893-917. DOI: 10.1002/ijc.25516

Fernandez SV, Bingham C, Fittipaldi P, Austin L, Palazzo J, Palmer G, Alpaugh K, Cristofanilli M. TP53 mutations detected in circulating tumor cells present in the blood of metastatic triple negative breast cancer patients. Breast Cancer Res. **2014** Oct 9;16(5):445. DOI: 10.1186/s13058-014-0445-3

Fidler IJ. Metastasis: quantitative analysis of distribution and fate of tumor emboli labeled with 125 I-5-iodo-2'-deoxyuridine. J Natl Cancer Inst. **1970** Oct;45(4):773-82. PMID: 5513503

Fidler IJ. Tumor heterogeneity and the biology of cancer invasion and metastasis. Cancer Res. **1978** Sep;38(9):2651-60. PMID: 354778

Fidler IJ. The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. Nat Rev Cancer. **2003** Jun;3(6):453-8. DOI:10.1038/nrc1098

Fidler IJ. Commentary on "Tumor Heterogeneity and the Biology of Cancer Invasion and Metastasis". Cancer Res. **2016** Jun 15;76(12):3441-2. DOI: 10.1158/0008-5472.CAN-16-1330

Fidler IJ and Kripke ML. Metastasis results from preexisting variant cells within a malignant tumor. Science. **1977** Aug 26;197(4306):893-5. PMID: 887927

Fina E, Cleris L, Reduzzi C, Morandi G, D'Aiuto F, Callari M, Daidone MG, Cappelletti V, Molecular characterization and biological role of circulating tumor cells: a window into metastasis biology from breast cancer xenograft models, 15th International Biennial Congress of the Metastasis Research Society, June 28 – July 1, **2014**, Heidelberg, Germany

Fina E, Callari M, Reduzzi C, D'Aiuto F, Mariani G, Generali D, Pierotti MA, Daidone MG, Cappelletti V. Gene expression profiling of circulating tumor cells in breast cancer. Clin Chem. **2015a** Jan;61(1):278-89. DOI: 10.1373/clinchem.2014.229476

Fina E, Reduzzi C, Motta R, Di Cosimo S, Bianchi G, Martinetti A, Wechsler J, Cappelletti V, Daidone MG. Did circulating tumor cells tell us all they could? The missed circulating tumor cell message in breast cancer. Int J Biol Markers. **2015b** Nov 11;30(4):e429-33. DOI: 10.5301/jbm.5000166

Fina E, Necchi A, Giannatempo P, Colecchia M, Raggi D, Daidone MG, and Cappelletti V.Clinical significance of early changes in circulating tumor cells from patients receiving first-line cisplatin-based chemotherapy for metastatic urothelial carcinoma. Bladder Cancer. **2016**; 2:395–403. DOI: 10.3233/BLC-160069

Fisher B, Dignam J, Wolmark N, DeCillis A, Emir B, Wickerham DL, Bryant J, Dimitrov NV, Abramson N, Atkins JN, Shibata H, Deschenes L, Margolese RG. Tamoxifen and chemotherapy for lymph node-negative, estrogen receptor-positive breast cancer. J Natl Cancer Inst. **1997** Nov 19;89(22):1673-82. DOI: 10.1093/jnci/89.22.1673

Fischer KR, Durrans A, Lee S, Sheng J, Li F, Wong ST, Choi H, El Rayes T, Ryu S, Troeger J, Schwabe RF, Vahdat LT, Altorki NK, Mittal V, Gao D. Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. Nature. **2015** Nov 26;527(7579):472-6. DOI: 10.1038/nature15748

Forster N, Saladi SV, van Bragt M, Sfondouris ME, Jones FE, Li Z, Ellisen LW. Basal cell signaling by p63 controls luminal progenitor function and lactation via NRG1. Dev Cell. **2014** Jan 27;28(2):147-60. DOI: 10.1016/j.devcel.2013.11.019

Fraker LD, Halter SA, Forbes JT. Growth inhibition by retinol of a human breast carcinoma cell line in vitro and in athymic mice. Cancer Res. **1984** Dec;44(12 Pt 1):5757-63. PMID: 6498837

Francia G, Emmenegger U, Lee CR, Shaked Y, Folkins C, Mossoba M, Medin JA, Man S, Zhu Z, Witte L, Kerbel RS. Long-term progression and therapeutic response of visceral metastatic disease non-invasively monitored in mouse urine using beta-human choriogonadotropin secreting tumor cell lines. Mol Cancer Ther. **2008** Oct;7(10):3452-9. DOI: 10.1158/1535-7163.MCT-08-0200

Franken B, de Groot MR, Mastboom WJ, Vermes I, van der Palen J, Tibbe AG, Terstappen LW. Circulating tumor cells, disease recurrence and survival in newly diagnosed breast cancer. Breast Cancer Res. **2012** Oct 22;14(5):R133. DOI: 10.1186/bcr3333

Friedl P and Wolf K. Plasticity of cell migration: a multiscale tuning model. J Cell Biol. **2010** Jan 11;188(1):11-9. DOI: 10.1083/jcb.200909003

Gao H, Chakraborty G, Lee-Lim AP, Mo Q, Decker M, Vonica A, Shen R, Brogi E, Brivanlou AH, Giancotti FG. The BMP inhibitor Coco reactivates breast cancer cells at lung metastatic sites. Cell. **2012** Aug 17;150(4):764-79. DOI: 10.1016/j.cell.2012.06.035

Garcia-Murillas I, Schiavon G, Weigelt B, Ng C, Hrebien S, Cutts RJ, Cheang M, Osin P, Nerurkar A, Kozarewa I, Garrido JA, Dowsett M, Reis-Filho JS, Smith IE, Turner NC. Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. Sci Transl Med. **2015** Aug 26;7(302):302ra133. DOI: 10.1126/scitranslmed.aab0021

Gay LJ and Felding-Habermann B. Contribution of platelets to tumour metastasis. Nat Rev Cancer. **2011** Feb;11(2):123-34. DOI: 10.1038/nrc3004

Gazzaniga P, Raimondi C, Gradilone A, Di Seri M, Longo F, Cortesi E, Frati L. Circulating tumor cells, colon cancer and bevacizumab: the meaning of zero. Ann Oncol. **2011** Aug;22(8):1929-30. DOI: 10.1093/annonc/mdr292

Geigl JB, Obenauf AC, Schwarzbraun T, Speicher MR. Defining 'chromosomal instability'. Trends Genet. **2008** Feb;24(2):64-9. DOI: 10.1016/j.tig.2007.11.006

Geyer CE, Forster J, Lindquist D, Chan S, Romieu CG, Pienkowski T, Jagiello-Gruszfeld A, Crown J, Chan A, Kaufman B, Skarlos D, Campone M, Davidson N, Berger M, Oliva C, Rubin SD, Stein S, Cameron D. Lapatinib plus capecitabine for HER2-positive advanced breast cancer. N Engl J Med. **2006** Dec 28;355(26):2733-43. DOI: 10.1056/NEJMoa064320

Ginestier C, Hur MH, Charafe-Jauffret E, Monville F, Dutcher J, Brown M, Jacquemier J, Viens P, Kleer CG, Liu S, Schott A, Hayes D, Birnbaum D, Wicha MS, Dontu G. ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. Cell Stem Cell. **2007** Nov;1(5):555-67. DOI: 10.1016/j.stem.2007.08.014

Glinsky VV, Glinsky GV, Glinskii OV, Huxley VH, Turk JR, Mossine VV, Deutscher SL, Pienta KJ, Quinn TP. Intravascular metastatic cancer cell homotypic aggregation at the sites of primary attachment to the endothelium. Cancer Res. **2003** Jul 1;63(13):3805-11. PMID: 12839977

Gomis RR, Alarcón C, Nadal C, Van Poznak C, Massagué J. C/EBPbeta at the core of the TGFbeta cytostatic response and its evasion in metastatic breast cancer cells. Cancer Cell. **2006** Sep;10(3):203-14. DOI: 10.1016/j.ccr.2006.07.019

Gordon D, Resio B, Pellman D. Causes and consequences of aneuploidy in cancer. Nat Rev Genet. **2012** Jan 24;13(3):189-203. DOI: 10.1038/nrg3123

Gralow JR, Burstein HJ, Wood W, Hortobagyi GN, Gianni L, von Minckwitz G, Buzdar AU, Smith IE, Symmans WF, Singh B, Winer EP Preoperative therapy in invasive breast cancer: pathologic assessment and systemic therapy issues in operable disease. J Clin Oncol. **2008** Feb 10;26(5):814-9. DOI: 10.1200/JCO.2007.15.3510

Graness A, Chwieralski CE, Reinhold D, Thim L, Hoffmann W. Protein kinase C and ERK activation are required for TFF-peptide-stimulated bronchial epithelial cell migration and tumor necrosis factor-alpha-induced interleukin-6 (IL-6) and IL-8 secretion. J Biol Chem. **2002** May 24;277(21):18440-6. DOI: 10.1074/jbc.M200468200

Granit RZ, Slyper M, Ben-Porath I. Axes of differentiation in breast cancer: untangling stemness, lineage identity, and the epithelial to mesenchymal transition. Wiley Interdiscip Rev Syst Biol Med. **2014** Jan-Feb;6(1):93-106. PMID: 24741710

Graus-Porta D, Beerli RR, Daly JM, Hynes NE. ErbB-2, the preferred heterodimerization partner of all ErbB receptors, is a mediator of lateral signaling. EMBO J. **1997** Apr 1;16(7):1647-55. DOI: 10.1093/emboj/16.7.1647

Gray JW, Collins C, Henderson IC, Isola J, Kallioniemi A, Kallioniemi OP, Nakamura H, Pinkel D, Stokke T, Tanner M, et al. Molecular cytogenetics of human breast cancer. Cold Spring Harb Symp Quant Biol. **1994**;59:645-52. PMID: 7587125

Greaves M and Maley CC. Clonal evolution in cancer. Nature. **2012** Jan 18;481(7381):306-13. DOI: 10.1038/nature10762

Gruvberger S, Ringnér M, Chen Y, Panavally S, Saal LH, Borg A, Fernö M, Peterson C, Meltzer PS. Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. Cancer Res. **2001** Aug 15;61(16):5979-84. PMID: 11507038

Guo W, Keckesova Z, Donaher JL, Shibue T, Tischler V, Reinhardt F, Itzkovitz S, Noske A, Zürrer-Härdi U, Bell G, Tam WL, Mani SA, van Oudenaarden A, Weinberg RA. Slug and Sox9 cooperatively determine the mammary stem cell state. Cell. **2012** Mar 2;148(5):1015-28. DOI: 10.1016/j.cell.2012.02.008

Gupta GP and Massagué J. Cancer metastasis: building a framework. Cell. **2006** Nov 17;127(4):679-95. DOI: 10.1016/j.cell.2006.11.001

Gupta GP, Perk J, Acharyya S, de Candia P, Mittal V, Todorova-Manova K, Gerald WL, Brogi E, Benezra R, Massagué J. ID genes mediate tumor reinitiation during breast cancer

lung metastasis. Proc Natl Acad Sci U S A. **2007** Dec 4;104(49):19506-11. DOI: 10.1073/pnas.0709185104

Gupta PB, Fillmore CM, Jiang G, Shapira SD, Tao K, Kuperwasser C, Lander ES. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. Cell. **2011** Aug 19;146(4):633-44. DOI: 10.1016/j.cell.2011.07.026

Guy CT, Cardiff RD, Muller WJ. Induction of mammary tumors by expression of polyomavirus middle T oncogene: a transgenic mouse model for metastatic disease. Mol Cell Biol. **1992** Mar;12(3):954-61. PMCID: PMC369527

Gyanchandani R, Lin Y, Lin HM, Cooper K, Normolle DP, Brufsky A, Fastuca M, Crosson W, Oesterreich S, Davidson NE, Bhargava R, Dabbs DJ, Lee AV. Intratumor Heterogeneity Affects Gene Expression Profile Test Prognostic Risk Stratification in Early Breast Cancer. Clin Cancer Res. **2016** Nov 1;22(21):5362-5369. DOI: 10.1158/1078-0432.CCR-15-2889

Hanahan D and Weinberg RA. Hallmarks of cancer: the next generation. Cell. **2011** Mar 4;144(5):646-74. DOI: 10.1016/j.cell.2011.02.013

Harari D and Yarden Y. Molecular mechanisms underlying ErbB2/HER2 action in breast cancer. Oncogene. **2000** Dec 11;19(53):6102-14. DOI: 10.1038/sj.onc.1203973

Harms JF and Welch DR. MDA-MB-435 human breast carcinoma metastasis to bone. Clin Exp Metastasis. **2003**;20(4):327-34. PMID: 12856720

Harris L, Fritsche H, Mennel R, Norton L, Ravdin P, Taube S, Somerfield MR, Hayes DF, Bast RC Jr; American Society of Clinical Oncology. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. J Clin Oncol. **2007** Nov 20;25(33):5287-312. DOI: 10.1200/JCO.2007.14.2364

Hart IR and Fidler IJ. Role of organ selectivity in the determination of metastatic patterns of B16 melanoma. Cancer Res. **1980** Jul;40(7):2281-7. PMID: 7388794

Hauser F, Poulsom R, Chinery R, Rogers LA, Hanby AM, Wright NA, Hoffmann W. hP1.B, a human P-domain peptide homologous with rat intestinal trefoil factor, is expressed also in the ulcer-associated cell lineage and the uterus. Proc Natl Acad Sci U S A. **1993** Aug 1;90(15):6961-5. PMCID: PMC47055

Hayes DF, Cristofanilli M, Budd GT, Ellis MJ, Stopeck A, Miller MC, Matera J, Allard WJ, Doyle GV, Terstappen LW. Circulating tumor cells at each follow-up time point during therapy of metastatic breast cancer patients predict progression-free and overall survival. Clin Cancer Res. **2006** Jul 15;12(14 Pt 1):4218-24. DOI: 10.1158/1078-0432.CCR-05-2821

He W, Wang H, Hartmann LC, Cheng JX, Low PS. In vivo quantitation of rare circulating tumor cells by multiphoton intravital flow cytometry. Proc Natl Acad Sci U S A. **2007** Jul 10;104(28):11760-5. DOI: 10.1073/pnas.0703875104

Helissey C, Berger F, Cottu P, Diéras V, Mignot L, Servois V, Bouleuc C, Asselain B, Pelissier S, Vaucher I, Pierga JY, Bidard FC. Circulating tumor cell thresholds and survival scores in advanced metastatic breast cancer: the observational step of the CirCe01 phase III trial. Cancer Lett. **2015** May 1;360(2):213-8. DOI: 10.1016/j.canlet.2015.02.010

Hennessy BT, Gonzalez-Angulo AM, Stemke-Hale K, Gilcrease MZ, Krishnamurthy S, Lee JS, Fridlyand J, Sahin A, Agarwal R, Joy C, Liu W, Stivers D, Baggerly K, Carey M, Lluch A, Monteagudo C, He X, Weigman V, Fan C, Palazzo J, Hortobagyi GN, Nolden LK, Wang NJ, Valero V, Gray JW, Perou CM, Mills GB. Characterization of a naturally occurring breast cancer subset enriched in epithelial-to-mesenchymal transition and stem cell characteristics. Cancer Res. **2009** May 15;69(10):4116-24. DOI: 10.1158/0008-5472.CAN-08-3441

Heppner GH, Miller FR, Shekhar PM. Nontransgenic models of breast cancer. Breast Cancer Res. **2000**;2(5):331-4. PMCID: PMC138654

Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, Rasmussen KE, Jones LP, Assefnia S, Chandrasekharan S, Backlund MG, Yin Y, Khramtsov AI, Bastein R, Quackenbush J, Glazer RI, Brown PH, Green JE, Kopelovich L, Furth PA, Palazzo JP, Olopade OI, Bernard PS, Churchill GA, Van Dyke T, Perou CM. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. Genome Biol. **2007**;8(5):R76. DOI: 10.1186/gb-2007-8-5-r76

Hodgkinson CL, Morrow CJ, Li Y, Metcalf RL, Rothwell DG, Trapani F, Polanski R, Burt DJ, Simpson KL, Morris K, Pepper SD, Nonaka D, Greystoke A, Kelly P, Bola B, Krebs MG, Antonello J, Ayub M, Faulkner S, Priest L, Carter L, Tate C, Miller CJ, Blackhall F, Brady G, Dive C. Tumorigenicity and genetic profiling of circulating tumor cells in small-cell lung cancer. Nat Med. **2014** Aug;20(8):897-903. DOI: 10.1038/nm.3600

Hoffmann W. Trefoil factor family (TFF) peptides and chemokine receptors: a promising relationship. J Med Chem. **2009** Nov 12;52(21):6505-10. DOI: 10.1021/jm9008136

Hofman VJ, Ilie MI, Bonnetaud C, Selva E, Long E, Molina T, Vignaud JM, Fléjou JF, Lantuejoul S, Piaton E, Butori C, Mourad N, Poudenx M, Bahadoran P, Sibon S, Guevara N, Santini J, Vénissac N, Mouroux J, Vielh P, Hofman PM. Cytopathologic detection of circulating tumor cells using the isolation by size of epithelial tumor cell method: promises and pitfalls. Am J Clin Pathol. **2011** Jan;135(1):146-56. DOI: 10.1309/AJCP9X8OZBEIQVVI

Hou JM, Krebs MG, Lancashire L, Sloane R, Backen A, Swain RK, Priest LJ, Greystoke A, Zhou C, Morris K, Ward T, Blackhall FH, Dive C. Clinical significance and molecular characteristics of circulating tumor cells and circulating tumor microemboli in patients with small-cell lung cancer. J Clin Oncol. **2012** Feb 10;30(5):525-32. DOI: 10.1200/JCO.2010.33.3716

Houssami N, Macaskill P, Balleine RL, Bilous M, Pegram MD. HER2 discordance between primary breast cancer and its paired metastasis: tumor biology or test artefact? Insights through meta-analysis. Breast Cancer Res Treat. **2011** Oct;129(3):659-74. DOI: 10.1007/s10549-011-1632-x

Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF, Livasy C, Carey LA, Reynolds E, Dressler L, Nobel A, Parker J, Ewend MG, Sawyer LR, Wu J, Liu Y, Nanda R, Tretiakova M, Ruiz Orrico A, Dreher D, Palazzo JP, Perreard L, Nelson E, Mone M, Hansen H, Mullins M, Quackenbush JF, Ellis MJ, Olopade OI, Bernard PS, Perou CM. The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics. **2006** Apr 27;7:96. DOI: 10.1186/1471-2164-7-96

Hudis CA. Trastuzumab--mechanism of action and use in clinical practice. N Engl J Med. **2007** Jul 5;357(1):39-51. DOI: 10.1056/NEJMra043186

Hüsemann Y, Geigl JB, Schubert F, Musiani P, Meyer M, Burghart E, Forni G, Eils R, Fehm T, Riethmüller G, Klein CA. Systemic spread is an early step in breast cancer. Cancer Cell. **2008** Jan;13(1):58-68. DOI: 10.1016/j.ccr.2007.12.003

Hynes NE and Dey JH. Potential for targeting the fibroblast growth factor receptors in breast cancer. Cancer Res. **2010** Jul 1;70(13):5199-202. DOI: 10.1158/0008-5472.CAN-10-0918

Ignatiadis M, Kallergi G, Ntoulia M, Perraki M, Apostolaki S, Kafousi M, Chlouverakis G, Stathopoulos E, Lianidou E, Georgoulias V, Mavroudis D. Prognostic value of the molecular detection of circulating tumor cells using a multimarker reverse transcription-PCR assay for cytokeratin 19, mammaglobin A, and HER2 in early breast cancer. Clin Cancer Res. **2008** May 1;14(9):2593-600. DOI: 10.1158/1078-0432.CCR-07-4758

Ignatiadis M, Rack B, Rothé F, Riethdorf S, Decraene C, Bonnefoi H, Dittrich C, Messina C, Beauvois M, Trapp E, Goulioti T, Tryfonidis K, Pantel K, Repollet M, Janni W, Piccart M, Sotiriou C, Litiere S, Pierga JY. Liquid biopsy-based clinical research in early breast cancer: The EORTC 90091-10093 Treat CTC trial. Eur J Cancer. **2016** Aug;63:97-104. DOI: 10.1016/j.ejca.2016.04.024

Inman JL, Robertson C, Mott JD, Bissell MJ. Mammary gland development: cell fate specification, stem cells and the microenvironment. Development. **2015** Mar 15;142(6):1028-42. DOI: 10.1242/dev.087643

Iorns E, Drews-Elger K, Ward TM, Dean S, Clarke J, Berry D, El Ashry D, Lippman M. A new mouse model for the study of human breast cancer metastasis. PLoS One. **2012**;7(10):e47995. DOI: 10.1371/journal.pone.0047995

Islam F, Qiao B, Smith RA, Gopalan V, Lam AK. Cancer stem cell: fundamental experimental pathological concepts and updates. Exp Mol Pathol. **2015** Apr;98(2):184-91. DOI: 10.1016/j.yexmp.2015.02.002

Iwakiri D and Podolsky DK. Keratinocyte growth factor promotes goblet cell differentiation through regulation of goblet cell silencer inhibitor. Gastroenterology. **2001** May;120(6):1372-80. PMID: 11313307

Janni WJ, Rack B, Terstappen LW, Pierga JY, Taran FA, Fehm T, Hall C, de Groot MR, Bidard FC, Friedl TW, Fasching PA, Brucker SY, Pantel K, Lucci A. Pooled Analysis of the Prognostic Relevance of Circulating Tumor Cells in Primary Breast Cancer. Clin Cancer Res. **2016** May 15;22(10):2583-93. DOI: 10.1158/1078-0432.CCR-15-1603

Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. CA Cancer J Clin. **2010** Sep-Oct;60(5):277-300. DOI: 10.3322/caac.20073

Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin. **2011** Mar-Apr;61(2):69-90. DOI: 10.3322/caac.20107

Jordan NV, Bardia A, Wittner BS, Benes C, Ligorio M, Zheng Y, Yu M, Sundaresan TK1, Licausi JA, Desai R, O'Keefe RM, Ebright RY, Boukhali M, Sil S, Onozato ML, Iafrate AJ,

Kapur R, Sgroi D, Ting DT, Toner M, Ramaswamy S, Haas W, Maheswaran S, Haber DA. HER2 expression identifies dynamic functional states within circulating breast cancer cells. Nature. **2016** Aug 24;537(7618):102-106. DOI: 10.1038/nature19328

Juratli MA, Sarimollaoglu M, Nedosekin DA, Melerzanov AV, Zharov VP, Galanzha EI. Dynamic Fluctuation of Circulating Tumor Cells during Cancer Progression. Cancers (Basel). **2014** Jan 15;6(1):128-42. DOI: 10.3390/cancers6010128

Kallioniemi OP, Kallioniemi A, Kurisu W, Thor A, Chen LC, Smith HS, Waldman FM, Pinkel D, Gray JW. ERBB2 amplification in breast cancer analyzed by fluorescence in situ hybridization. Proc Natl Acad Sci U S A. **1992** Jun 15;89(12):5321-5. PMCID: PMC49283

Kang Y, Siegel PM, Shu W, Drobnjak M, Kakonen SM, Cordón-Cardo C, Guise TA, Massagué J. A multigenic program mediating breast cancer metastasis to bone. Cancer Cell. **2003** Jun;3(6):537-49. DOI: 10.1016/S1535-6108(03)00132-6

Kang Y, He W, Tulley S, Gupta GP, Serganova I, Chen CR, Manova-Todorova K, Blasberg R, Gerald WL, Massagué J. Breast cancer bone metastasis mediated by the Smad tumor suppressor pathway. Proc Natl Acad Sci U S A. **2005** Sep 27;102(39):13909-14. DOI: 10.1073/pnas.0506517102

Kang Y, Pantel K. Tumor cell dissemination: emerging biological insights from animal models and cancer patients. Cancer Cell. **2013** May 13;23(5):573-81. DOI: 10.1016/j.ccr.2013.04.017

Kannan N, Kang J, Kong X, Tang J, Perry JK, Mohankumar KM, Miller LD, Liu ET, Mertani HC, Zhu T, Grandison PM, Liu DX, Lobie PE. Trefoil factor 3 is oncogenic and mediates anti-estrogen resistance in human mammary carcinoma. Neoplasia. **2010** Dec;12(12):1041-53. PMCID: PMC3003139

Kanwar N, Hu P, Bedard P, Clemons M, McCready D, Done SJ. Identification of genomic signatures in circulating tumor cells from breast cancer. Int J Cancer. **2015** Jul 15;137(2):332-44. DOI: 10.1002/ijc.29399

Kato S, Endoh H, Masuhiro Y, Kitamoto T, Uchiyama S, Sasaki H, Masushige S, Gotoh Y, Nishida E, Kawashima H, Metzger D, Chambon P. Activation of the estrogen receptor through phosphorylation by mitogen-activated protein kinase. Science. **1995** Dec 1;270(5241):1491-4. PMID: 7491495

Kats-Ugurlu G, Roodink I, de Weijert M, Tiemessen D, Maass C, Verrijp K, van der Laak J, de Waal R, Mulders P, Oosterwijk E, Leenders W. Circulating tumour tissue fragments in patients with pulmonary metastasis of clear cell renal cell carcinoma. J Pathol. **2009** Nov;219(3):287-93. DOI: 10.1002/path.2613

Kaufmann M, Hortobagyi GN, Goldhirsch A, Scholl S, Makris A, Valagussa P, Blohmer JU, Eiermann W, Jackesz R, Jonat W, Lebeau A, Loibl S, Miller W, Seeber S, Semiglazov V, Smith R, Souchon R, Stearns V, Untch M, von Minckwitz G. Recommendations from an international expert panel on the use of neoadjuvant (primary) systemic treatment of operable breast cancer: an update. J Clin Oncol. **2006** Apr 20;24(12):1940-9. DOI: 10.1200/JCO.2005.02.6187

Kennecke H, Yerushalmi R, Woods R, Cheang MC, Voduc D, Speers CH, Nielsen TO, Gelmon K. Metastatic behavior of breast cancer subtypes. J Clin Oncol. **2010** Jul 10;28(20):3271-7. DOI: 10.1200/JCO.2009.25.9820

Key J, Hodgson S, Omar RZ, Jensen TK, Thompson SG, Boobis AR, Davies DS, Elliott P. Meta-analysis of studies of alcohol and breast cancer with consideration of the methodological issues. Cancer Causes Control. **2006** Aug;17(6):759-70. DOI: 10.1007/s10552-006-0011-0

Key TJ, Verkasalo PK, Banks E. Epidemiology of breast cancer. Lancet Oncol. **2001** Mar;2(3):133-40. DOI: 10.1016/S1470-2045(00)00254-0

Khanna C and Hunter K. Modeling metastasis in vivo. Carcinogenesis. **2005** Mar;26(3):513-23. Epub 2004 Sep 9. DOI: 10.1093/carcin/bgh261

Kim MY, Oskarsson T, Acharyya S, Nguyen DX, Zhang XH, Norton L, Massagué J.Cell. Tumor self-seeding by circulating cancer cells. **2009** Dec 24;139(7):1315-26. DOI: 10.1016/j.cell.2009.11.025

Kim SJ, Masago A, Tamaki Y, Akazawa K, Tsukamoto F, Sato J, Ozawa T, Tsujino Y, Noguchi S. A novel approach using telomerase-specific replication-selective adenovirus for detection of circulating tumor cells in breast cancer patients. Breast Cancer Res Treat. **2011** Aug;128(3):765-73. DOI: 10.1007/s10549-011-1603-2

Kim YN, Koo KH, Sung JY, Yun UJ, Kim H. Anoikis resistance: an essential prerequisite for tumor metastasis. Int J Cell Biol. **2012**;2012:306879. DOI: 10.1155/2012/306879

Kjellev S. The trefoil factor family - small peptides with multiple functionalities. Cell Mol Life Sci. **2009** Apr;66(8):1350-69. DOI: 10.1007/s00018-008-8646-5.

Klein CA. Parallel progression of primary tumours and metastases. Nat Rev Cancer. **2009** Apr;9(4):302-12. DOI: 10.1038/nrc2627

Klein G. Foulds' dangerous idea revisited: the multistep development of tumors 40 years later. Adv Cancer Res. **1998**;72:1-23. PMID: 9338072

Klevebring D, Rosin G, Ma R, Lindberg J, Czene K, Kere J, Fredriksson I, Bergh J, Hartman J. Sequencing of breast cancer stem cell populations indicates a dynamic conversion between differentiation states in vivo. Breast Cancer Res. **2014** Jul 6;16(4):R72. DOI: 10.1186/bcr3687

Konecny G, Pauletti G, Pegram M, Untch M, Dandekar S, Aguilar Z, Wilson C, Rong HM, Bauerfeind I, Felber M, Wang HJ, Beryt M, Seshadri R, Hepp H, Slamon DJ. Quantitative association between HER-2/neu and steroid hormone receptors in hormone receptor-positive primary breast cancer. J Natl Cancer Inst. **2003** Jan 15;95(2):142-53. DOI: 10.1093/jnci/95.2.142

Kordon EC and Smith GH. An entire functional mammary gland may comprise the progeny from a single cell. Development. **1998** May;125(10):1921-30. PMID: 9550724

Korkaya H and Wicha MS. HER2 and breast cancer stem cells: more than meets the eye. Cancer Res. **2013** Jun 15;73(12):3489-93. DOI: 10.1158/0008-5472.CAN-13-0260

Kowalewska M, Chechlińska M, Markowicz S, Kober P, Nowak R. The relevance of RT-PCR detection of disseminated tumour cells is hampered by the expression of markers regarded as tumour-specific in activated lymphocytes. Eur J Cancer. **2006** Nov;42(16):2671-4. DOI: 10.1016/j.ejca.2006.05.036

Kozlow W and Guise TA. Breast cancer metastasis to bone: mechanisms of osteolysis and implications for therapy. J Mammary Gland Biol Neoplasia. **2005** Apr;10(2):169-80. DOI: 10.1007/s10911-005-5399-8

Kretzschmar K and Watt FM. Lineage tracing. Cell. **2012** Jan 20;148(1-2):33-45. DOI: 10.1016/j.cell.2012.01.002

Kripke ML, Gruys E, Fidler IJ. Metastatic heterogeneity of cells from an ultraviolet light-induced murine fibrosarcoma of recent origin. Cancer Res. **1978** Sep;38(9):2962-7. PMID: 679204

Labelle M, Begum S, Hynes RO. Direct signaling between platelets and cancer cells induces an epithelial-mesenchymal-like transition and promotes metastasis. Cancer Cell. **2011** Nov 15;20(5):576-90. DOI: 10.1016/j.ccr.2011.09.009

Lacroix M and Leclercq G. Relevance of breast cancer cell lines as models for breast tumours: an update. Breast Cancer Res Treat. **2004** Feb;83(3):249-89. DOI: 10.1023/B:BREA.0000014042.54925.cc

Lagadec C, Vlashi E, Della Donna L, Meng Y, Dekmezian C, Kim K, Pajonk F.Breast Cancer Res. Survival and self-renewing capacity of breast cancer initiating cells during fractionated radiation treatment. **2010**;12(1):R13. DOI: 10.1186/bcr2479

Lakhani S, Ellis I, Schnitt S, et al. **2012**.WHO Classification of Tumours of the Breast. 4th. Lyon: IARC Press

Lamb R, Ablett MP, Spence K, Landberg G, Sims AH, Clarke RB. Wnt pathway activity in breast cancer sub-types and stem-like cells. PLoS One. **2013** Jul 4;8(7):e67811. DOI: 10.1371/journal.pone.0067811

Landis MD, Lehmann BD, Pietenpol JA, Chang JC. Patient-derived breast tumor xenografts facilitating personalized cancer therapy. Breast Cancer Res. **2013** Jan 22;15(1):201. DOI: 10.1186/bcr3355

Lasa A, Garcia A, Alonso C, Millet P, Cornet M, Ramón y Cajal T, Baiget M, Barnadas A Molecular detection of peripheral blood breast cancer mRNA transcripts as a surrogate biomarker for circulating tumor cells. PLoS One. **2013** Sep 18;8(9):e74079. DOI: 10.1371/journal.pone.0074079

Lau WH, Pandey V, Kong X, Wang XN, Wu Z, Zhu T, Lobie PE. Trefoil Factor-3 (TFF3) Stimulates De Novo Angiogenesis in Mammary Carcinoma both Directly and Indirectly via IL-8/CXCR2. PLoS One. **2015** Nov 11;10(11):e0141947. DOI: 10.1371/journal.pone.0141947

Lawson DA, Bhakta NR, Kessenbrock K, Prummel KD, Yu Y, Takai K, Zhou A, Eyob H, Balakrishnan S, Wang CY, Yaswen P, Goga A, Werb Z. Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells. Nature. **2015** Oct 1;526(7571):131-5. DOI: 10.1038/nature15260

Lazova R, Laberge GS, Duvall E, Spoelstra N, Klump V, Sznol M, Cooper D, Spritz RA, Chang JT, Pawelek JM. A Melanoma Brain Metastasis with a Donor-Patient Hybrid Genome following Bone Marrow Transplantation: First Evidence for Fusion in Human Cancer. PLoS One. **2013** Jun 26;8(6):e66731. DOI: 10.1371/journal.pone.0066731

LeBleu VS, O'Connell JT, Gonzalez Herrera KN, Wikman H, Pantel K, Haigis MC, de Carvalho FM, Damascena A, Domingos Chinen LT, Rocha RM, Asara JM, Kalluri R. PGC-1α mediates mitochondrial biogenesis and oxidative phosphorylation in cancer cells to promote metastasis. Nat Cell Biol. **2014** Oct;16(10):992-1003, 1-15. DOI: 10.1038/ncb3039

Lee GY, Kenny PA, Lee EH, Bissell MJ. Three-dimensional culture models of normal and malignant breast epithelial cells. Nat Methods. **2007** Apr;4(4):359-65. DOI: 10.1038/nmeth1015

Lee HE, Kim JH, Kim YJ, Choi SY, Kim SW, Kang E, Chung IY, Kim IA, Kim EJ, Choi Y, Ryu HS, Park SY. An increase in cancer stem cell population after primary systemic therapy is a poor prognostic factor in breast cancer. Br J Cancer. **2011** May 24;104(11):1730-8. DOI: 10.1038/bjc.2011.159

Lee JY, Park K, Lee E, Ahn T, Jung HH, Lim SH, Hong M, Do IG, Cho EY, Kim DH, Kim JY, Ahn JS, Im YH, Park YH. Gene Expression Profiling of Breast Cancer Brain Metastasis. Sci Rep. **2016** Jun 24;6:28623. DOI: 10.1038/srep28623

Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, Pietenpol JA. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest. **2011** Jul;121(7):2750-67. DOI: 10.1172/JCI45014

Leis O, Eguiara A, Lopez-Arribillaga E, Alberdi MJ, Hernandez-Garcia S, Elorriaga K, Pandiella A, Rezola R, Martin AG. Sox2 expression in breast tumours and activation in breast cancer stem cells. Oncogene. **2012** Mar 15;31(11):1354-65. DOI: 10.1038/onc.2011.338

Lelekakis M, Moseley JM, Martin TJ, Hards D, Williams E, Ho P, Lowen D, Javni J, Miller FR, Slavin J, Anderson RL. A novel orthotopic model of breast cancer metastasis to bone. Clin Exp Metastasis. **1999** Mar;17(2):163-70. PMID: 10411109

Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. Nature. **1998** Dec 17;396(6712):643-9.

Li HZ, Yi TB, Wu ZY. Suspension culture combined with chemotherapeutic agents for sorting of breast cancer stem cells. BMC Cancer. **2008** May 14;8:135. DOI: 10.1186/1471-2407-8-135

Li X, Lewis MT, Huang J, Gutierrez C, Osborne CK, Wu MF, Hilsenbeck SG, Pavlick A, Zhang X, Chamness GC, Wong H, Rosen J, Chang JC. Intrinsic resistance of tumorigenic breast cancer cells to chemotherapy. J Natl Cancer Inst. **2008** May 7;100(9):672-9. DOI: 10.1093/jnci/djn123

Lim E, Vaillant F, Wu D, Forrest NC, Pal B, Hart AH, Asselin-Labat ML, Gyorki DE, Ward T, Partanen A, Feleppa F, Huschtscha LI, Thorne HJ; kConFab, Fox SB, Yan M, French JD, Brown MA, Smyth GK, Visvader JE, Lindeman GJ. Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. Nat Med. **2009** Aug;15(8):907-13. DOI: 10.1038/nm.2000

Lin EY, Jones JG, Li P, Zhu L, Whitney KD, Muller WJ, Pollard JW. Progression to malignancy in the polyoma middle T oncoprotein mouse breast cancer model provides a reliable model for human diseases. Am J Pathol. **2003** Nov;163(5):2113-26. DOI: 10.1016/S0002-9440(10)63568-7

Liu D, el-Hariry I, Karayiannakis AJ, Wilding J, Chinery R, Kmiot W, McCrea PD, Gullick WJ, Pignatelli M. Phosphorylation of beta-catenin and epidermal growth factor receptor by intestinal trefoil factor. Lab Invest. **1997** Dec;77(6):557-63. PMID: 9426392

Liu R, Wang X, Chen GY, Dalerba P, Gurney A, Hoey T, Sherlock G, Lewicki J, Shedden K, Clarke MF. The prognostic role of a gene signature from tumorigenic breast-cancer cells. N Engl J Med. **2007** Jan 18;356(3):217-26. DOI: 10.1056/NEJMoa063994

Liu S, Ginestier C, Ou SJ, Clouthier SG, Patel SH, Monville F, Korkaya H, Heath A, Dutcher J, Kleer CG, Jung Y, Dontu G, Taichman R, Wicha MS. Breast cancer stem cells are regulated by mesenchymal stem cells through cytokine networks. Cancer Res. **2011** Jan 15;71(2):614-24. DOI: 10.1158/0008-5472.CAN-10-0538

Liu S, Cong Y, Wang D, Sun Y, Deng L, Liu Y, Martin-Trevino R, Shang L, McDermott SP, Landis MD, Hong S, Adams A, D'Angelo R, Ginestier C, Charafe-Jauffret E, Clouthier SG, Birnbaum D, Wong ST, Zhan M, Chang JC, Wicha MS. Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. Stem Cell Reports. **2013** Dec 27;2(1):78-91. DOI: 10.1016/j.stemcr.2013.11.009

Livraghi L and Garber JE. PARP inhibitors in the management of breast cancer: current data and future prospects. BMC Med. **2015** Aug 13;13:188. DOI: 10.1186/s12916-015-0425-1

Loi S, Michiels S, Salgado R, Sirtaine N, Jose V, Fumagalli D, Kellokumpu-Lehtinen PL, Bono P, Kataja V, Desmedt C, Piccart MJ, Loibl S, Denkert C, Smyth MJ, Joensuu H, Sotiriou C. Tumor infiltrating lymphocytes are prognostic in triple negative breast cancer and predictive for trastuzumab benefit in early breast cancer: results from the FinHER trial. Ann Oncol. **2014** Aug;25(8):1544-50. DOI: 10.1093/annonc/mdu112

Lu J, Fan T, Zhao Q, Zeng W, Zaslavsky E, Chen JJ, Frohman MA, Golightly MG, Madajewicz S, Chen WT. Isolation of circulating epithelial and tumor progenitor cells with an invasive phenotype from breast cancer patients. Int J Cancer. **2010** Feb 1;126(3):669-83. DOI: 10.1002/ijc.24814

Lucci A, Hall CS, Lodhi AK, Bhattacharyya A, Anderson AE, Xiao L, Bedrosian I, Kuerer HM, Krishnamurthy S. Circulating tumour cells in non-metastatic breast cancer: a prospective study. Lancet Oncol. **2012** Jul;13(7):688-95. DOI: 10.1016/S1470-2045(12)70209-7

Luo X, Mitra D, Sullivan RJ, Wittner BS, Kimura AM, Pan S, Hoang MP, Brannigan BW, Lawrence DP, Flaherty KT, Sequist LV, McMahon M, Bosenberg MW, Stott SL, Ting DT, Ramaswamy S, Toner M, Fisher DE, Maheswaran S, Haber DA. Isolation and molecular characterization of circulating melanoma cells. Cell Rep. **2014** May 8;7(3):645-53. DOI: 10.1016/j.celrep.2014.03.039

Luzzi KJ, MacDonald IC, Schmidt EE, Kerkvliet N, Morris VL, Chambers AF, Groom AC. Multistep nature of metastatic inefficiency: dormancy of solitary cells after successful extravasation and limited survival of early micrometastases. Am J Pathol. **1998** Sep;153(3):865-73. DOI: 10.1016/S0002-9440(10)65628-3

Ma XJ, Salunga R, Tuggle JT, Gaudet J, Enright E, McQuary P, Payette T, Pistone M, Stecker K, Zhang BM, Zhou YX, Varnholt H, Smith B, Gadd M, Chatfield E, Kessler J, Baer TM, Erlander MG, Sgroi DC. Gene expression profiles of human breast cancer progression. Proc Natl Acad Sci U S A. **2003** May 13;100(10):5974-9. DOI: 10.1073/pnas.0931261100

Macias H and Hinck L. Mammary gland development. Wiley Interdiscip Rev Dev Biol. **2012** Jul-Aug;1(4):533-57. DOI: 10.1002/wdev.35

Madsen J, Sorensen GL, Nielsen O, Tornøe I, Thim L, Fenger C, Mollenhauer J, Holmskov U. A variant form of the human deleted in malignant brain tumor 1 (DMBT1) gene shows increased expression in inflammatory bowel diseases and interacts with dimeric trefoil factor 3 (TFF3). PLoS One. **2013** May 15;8(5):e64441. DOI: 10.1371/journal.pone.0064441

Magbanua MJ, Melisko M, Roy R, Sosa EV, Hauranieh L, Kablanian A, Eisenbud LE, Ryazantsev A, Au A, Scott JH, Park JW. Molecular profiling of tumor cells in cerebrospinal fluid and matched primary tumors from metastatic breast cancer patients with leptomeningeal carcinomatosis. Cancer Res. **2013** Dec 1;73(23):7134-43. DOI: 10.1158/0008-5472.CAN-13-2051

Maglione JE, Moghanaki D, Young LJ, Manner CK, Ellies LG, Joseph SO, Nicholson B, Cardiff RD, MacLeod CL. Transgenic Polyoma middle-T mice model premalignant mammary disease. Cancer Res. **2001** Nov 15;61(22):8298-305. PMID: 11719463

Magnifico A, Albano L, Campaner S, Delia D, Castiglioni F, Gasparini P, Sozzi G, Fontanella E, Menard S, Tagliabue E. Tumor-initiating cells of HER2-positive carcinoma cell lines express the highest oncoprotein levels and are sensitive to trastuzumab. Clin Cancer Res. **2009** Mar 15;15(6):2010-21. DOI: 10.1158/1078-0432.CCR-08-1327

Malladi S, Macalinao DG, Jin X, He L, Basnet H, Zou Y, de Stanchina E, Massagué J. Metastatic Latency and Immune Evasion through Autocrine Inhibition of WNT. Cell. **2016** Mar 24;165(1):45-60. DOI: 10.1016/j.cell.2016.02.025

Mani SA, Guo W, Liao MJ, Eaton EN, Ayyanan A, Zhou AY, Brooks M, Reinhard F, Zhang CC, Shipitsin M, Campbell LL, Polyak K, Brisken C, Yang J, Weinberg RA. The epithelial-

mesenchymal transition generates cells with properties of stem cells. Cell. **2008** May 16;133(4):704-15. DOI: 10.1016/j.cell.2008.03.027

Mansour EG, Gray R, Shatila AH, Osborne CK, Tormey DC, Gilchrist KW, Cooper MR, Falkson G. Efficacy of adjuvant chemotherapy in high-risk node-negative breast cancer. An intergroup study. N Engl J Med. **1989** Feb 23;320(8):485-90. DOI: 10.1056/NEJM198902233200803

Marangoni E, Vincent-Salomon A, Auger N, Degeorges A, Assayag F, de Cremoux P, de Plater L, Guyader C, De Pinieux G, Judde JG, Rebucci M, Tran-Perennou C, Sastre-Garau X, Sigal-Zafrani B, Delattre O, Diéras V, Poupon MF. A new model of patient tumor-derived breast cancer xenografts for preclinical assays. Clin Cancer Res. **2007** Jul 1;13(13):3989-98. DOI: 10.1158/1078-0432.CCR-07-0078

Marino M, Galluzzo P, Ascenzi P. Estrogen signaling multiple pathways to impact gene transcription. Curr Genomics. **2006**;7(8):497-508. PMCID: PMC2269003

Markou A, Farkona S, Schiza C, Efstathiou T, Kounelis S, Malamos N, Georgoulias V, Lianidou E. PIK3CA mutational status in circulating tumor cells can change during disease recurrence or progression in patients with breast cancer. Clin Cancer Res. **2014** Nov 15;20(22):5823-34. DOI: 10.1158/1078-0432.CCR-14-0149

Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? Nat Rev Cancer. **2012** Apr 19;12(5):323-34. DOI: 10.1038/nrc3261

Masiakowski P, Breathnach R, Bloch J, Gannon F, Krust A, Chambon P. Cloning of cDNA sequences of hormone-regulated genes from the MCF-7 human breast cancer cell line. Nucleic Acids Res. **1982** Dec 20;10(24):7895-903. PMCID: PMC327057

Massagué J. TGFβ signalling in context. Nat Rev Mol Cell Biol. **2012** Oct;13(10):616-30. DOI: 10.1038/nrm3434

Massagué J and Obenauf AC. Metastatic colonization by circulating tumour cells. Nature. **2016** Jan 21;529(7586):298-306. DOI: 10.1038/nature17038

Mastracci TL, Shadeo A, Colby SM, Tuck AB, O'Malley FP, Bull SB, Lam WL, Andrulis IL. Genomic alterations in lobular neoplasia: a microarray comparative genomic hybridization signature for early neoplastic proliferationin the breast. Genes Chromosomes Cancer. **2006** Nov;45(11):1007-17. DOI: 10.1002/gcc.20368

Mavaddat N, Antoniou AC, Easton DF, Garcia-Closas M. Genetic susceptibility to breast cancer. Mol Oncol. **2010** Jun;4(3):174-91. DOI: 10.1016/j.molonc.2010.04.011

May CD, Sphyris N, Evans KW, Werden SJ, Guo W, Mani SA. Epithelial-mesenchymal transition and cancer stem cells: a dangerously dynamic duo in breast cancer progression. Breast Cancer Res. **2011** Feb 8;13(1):202. DOI: 10.1186/bcr2789

May FE and Westley BR. Expression of human intestinal trefoil factor in malignant cells and its regulation by oestrogen in breast cancer cells. J Pathol. **1997** Aug;182(4):404-13. DOI: 10.1002/(SICI)1096-9896(199708)182:4<404::AID-PATH875>3.0.CO;2-0

May FE and Westley BR. TFF3 is a valuable predictive biomarker of endocrine response in metastatic breast cancer. Endocr Relat Cancer. **2015** Jun;22(3):465-79. DOI: 10.1530/ERC-15-0129

May FE, Church ST, Major S, Westley BR. The closely related estrogen-regulated trefoil proteins TFF1 and TFF3 have markedly different hydrodynamic properties, overall charge, and distribution of surface charge. Biochemistry. **2003** Jul 15;42(27):8250-9. DOI: 10.1021/bi030025l

McCall MN, Bolstad BM, Irizarry RA. Frozen robust multiarray analysis (fRMA). Biostatistics. **2010** Apr;11(2):242-53. DOI: 10.1093/biostatistics/kxp059

McClatchey AI. Modeling metastasis in the mouse. Oncogene. **1999** Sep 20;18(38):5334-9. DOI: 10.1038/sj.onc.1203086

McGranahan N, Burrell RA, Endesfelder D, Novelli MR, Swanton C. Cancer chromosomal instability: therapeutic and diagnostic challenges. EMBO Rep. **2012** Jun 1;13(6):528-38. DOI: 10.1038/embor.2012.61

Medoro G, et al. A Lab-on-a-Chip for Cell Detection and Manipulation. **2003**. IEEE Sensor, 3(3):317–325

Mego M, Mani SA, Cristofanilli M. Molecular mechanisms of metastasis in breast cancer--clinical applications. Nat Rev Clin Oncol. **2010** Dec;7(12):693-701. DOI: 10.1038/nrclinonc.2010.171

Meng S, Tripathy D, Shete S, Ashfaq R, Haley B, Perkins S, Beitsch P, Khan A, Euhus D, Osborne C, Frenkel E, Hoover S, Leitch M, Clifford E, Vitetta E, Morrison L, Herlyn D, Terstappen LW, Fleming T, Fehm T, Tucker T, Lane N, Wang J, Uhr J. HER-2 gene amplification can be acquired as breast cancer progresses. Proc Natl Acad Sci U S A. **2004a** Jun 22;101(25):9393-8. DOI: 10.1073/pnas.0402993101

Meng S, Tripathy D, Frenkel EP, Shete S, Naftalis EZ, Huth JF, Beitsch PD, Leitch M, Hoover S, Euhus D, Haley B, Morrison L, Fleming TP, Herlyn D, Terstappen LW, Fehm T, Tucker TF, Lane N, Wang J, Uhr JW. Circulating tumor cells in patients with breast cancer dormancy. Clin Cancer Res. **2004b** Dec 15;10(24):8152-62. DOI: 10.1158/1078-0432.CCR-04-1110

Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, Liu Q, Cochran C, Bennett LM, Ding W, et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science. **1994** Oct 7;266(5182):66-71. PMID: 7545954

Miller FR, Miller BE, Heppner GH. Characterization of metastatic heterogeneity among subpopulations of a single mouse mammary tumor: heterogeneity in phenotypic stability. Invasion Metastasis. **1983**;3(1):22-31. PMID: 6677618

Minn AJ, Gupta GP, Siegel PM, Bos PD, Shu W, Giri DD, Viale A, Olshen AB, Gerald WL, Massagué J. Genes that mediate breast cancer metastasis to lung. Nature. **2005** Jul 28;436(7050):518-24. DOI: 10.1038/nature03799

Mitchell MJ and King MR. Computational and experimental models of cancer cell response to fluid shear stress. Front Oncol. **2013** Mar 5;3:44. DOI: 10.3389/fonc.2013.00044

Moasser MM. The oncogene HER2: its signaling and transforming functions and its role in human cancer pathogenesis. Oncogene. **2007** Oct 4;26(45):6469-87. DOI:10.1038/sj.onc.1210477

Mohme M, Riethdorf S, Pantel K. Circulating and disseminated tumour cells - mechanisms of immune surveillance and escape. Nat Rev Clin Oncol. **2016** Sep 20. DOI: 10.1038/nrclinonc.2016.144

Molnar B, Ladanyi A, Tanko L, Sréter L, Tulassay Z. Circulating tumor cell clusters in the peripheral blood of colorectal cancer patients. Clin Cancer Res. **2001** Dec;7(12):4080-5. PMID: 11751505

Moore KL, Persod TVN, Torchia MG. The Developing Human. Clinically Oriented Embryology. 9th ed. **2013**. Philadelphia, PA: Elsevier Saunders

Morandi L, Marucci G, Foschini MP, Cattani MG, Pession A, Riva C, Eusebi V. Genetic similarities and differences between lobular in situ neoplasia (LN) and invasive lobular carcinoma of the breast. Virchows Arch. **2006** Jul;449(1):14-23. DOI: 10.1007/s00428-006-0192-7

Morley P, Whitfield JF, Vanderhyden BC, Tsang BK, Schwartz JL. A new, nongenomic estrogen action: the rapid release of intracellular calcium. Endocrinology. **1992** Sep;131(3):1305-12. DOI: 10.1210/endo.131.3.1505465

Mostert B, Sieuwerts AM, Kraan J, Bolt-de Vries J, van der Spoel P, van Galen A, Peeters DJ, Dirix LY, Seynaeve CM, Jager A, de Jongh FE, Hamberg P, Stouthard JM, Kehrer DF, Look MP, Smid M, Gratama JW, Foekens JA, Martens JW, Sleijfer S. Gene expression profiles in circulating tumor cells to predict prognosis in metastatic breast cancer patients. Ann Oncol. **2015** Mar;26(3):510-6. DOI: 10.1093/annonc/mdu557

Mu Z, Benali-Furet N, Uzan G, Znaty A, Ye Z, Paolillo C, Wang C, Austin L, Rossi G, Fortina P, Yang H, Cristofanilli M. Detection and Characterization of Circulating Tumor Associated Cells in Metastatic Breast Cancer. Int J Mol Sci. **2016** Sep 30;17(10). DOI: 10.3390/ijms17101665

Muenst S, Soysal SD, Gao F, Obermann EC, Oertli D, Gillanders WE. The presence of programmed death 1 (PD-1)-positive tumor-infiltrating lymphocytes is associated with poor prognosis in human breast cancer. Breast Cancer Res Treat. **2013** Jun;139(3):667-76. DOI: 10.1007/s10549-013-2581-3

Muller WJ, Sinn E, Pattengale PK, Wallace R, Leder P. Single-step induction of mammary adenocarcinoma in transgenic mice bearing the activated c-neu oncogene. Cell. **1988** Jul 1;54(1):105-15. PMID: 2898299

Müller V, Riethdorf S, Rack B, Janni W, Fasching PA, Solomayer E, Aktas B, Kasimir-Bauer S, Pantel K, Fehm T; DETECT study group. Prognostic impact of circulating tumor cells assessed with the CellSearch System™ and AdnaTest Breast™ in metastatic breast cancer

patients: the DETECT study. Breast Cancer Res. **2012** Aug 15;14(4):R118. DOI: 10.1186/bcr3243

Munoz R, Man S, Shaked Y, Lee CR, Wong J, Francia G, Kerbel RS. Highly efficacious nontoxic preclinical treatment for advanced metastatic breast cancer using combination oral UFT-cyclophosphamide metronomic chemotherapy. Cancer Res. **2006** Apr 1;66(7):3386-91. DOI: 10.1158/0008-5472.CAN-05-4411

Musella V, Pietrantonio F, Di Buduo E, Iacovelli R, Martinetti A, Sottotetti E, Bossi I, Maggi C, Di Bartolomeo M, de Braud F, Daidone MG, Cappelletti V. Circulating tumor cells as a longitudinal biomarker in patients with advanced chemorefractory, RAS-BRAF wild-type colorectal cancer receiving cetuximab or panitumumab. Int J Cancer. **2015** Sep 15;137(6):1467-74. DOI: 10.1002/ijc.29493

Nagaraja GM, Othman M, Fox BP, Alsaber R, Pellegrino CM, Zeng Y, Khanna R, Tamburini P, Swaroop A, Kandpal RP. Gene expression signatures and biomarkers of noninvasive and invasive breast cancer cells: comprehensive profiles by representational difference analysis, microarrays and proteomics. Oncogene. **2006** Apr 13;25(16):2328-38. DOI: 10.1038/sj.onc.1209265

Nagrath S, Sequist LV, Maheswaran S, Bell DW, Irimia D, Ulkus L, Smith MR, Kwak EL, Digumarthy S, Muzikansky A, Ryan P, Balis UJ, Tompkins RG, Haber DA, Toner M. Isolation of rare circulating tumour cells in cancer patients by microchip technology. Nature. **2007** Dec 20;450(7173):1235-9. DOI: 10.1038/nature06385

Nanda R, Chow LQ2, Dees EC2, Berger R2, Gupta S2, Geva R2, Pusztai L2, Pathiraja K2, Aktan G2, Cheng JD2, Karantza V2, Buisseret L2. Pembrolizumab in Patients With Advanced Triple-Negative Breast Cancer: Phase Ib KEYNOTE-012 Study. J Clin Oncol. **2016** Jul 20;34(21):2460-7. DOI: 10.1200/JCO.2015.64.8931

Nanni P, Nicoletti G, Palladini A, Croci S, Murgo A, Ianzano ML, Grosso V, Stivani V, Antognoli A, Lamolinara A, Landuzzi L, di Tomaso E, Iezzi M, De Giovanni C, Lollini PL. Multiorgan metastasis of human HER-2+ breast cancer in Rag2-/-;Il2rg-/- mice and treatment with PI3K inhibitor. PLoS One. **2012**;7(6):e39626. DOI: 10.1371/journal.pone.0039626

Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, Clark L, Bayani N, Coppe JP, Tong F, Speed T, Spellman PT, DeVries S, Lapuk A, Wang NJ, Kuo WL, Stilwell JL, Pinkel D, Albertson DG, Waldman FM, McCormick F, Dickson RB, Johnson MD, Lippman M, Ethier S, Gazdar A, Gray JW. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell. **2006** Dec;10(6):515-27. DOI: 10.1016/j.ccr.2006.10.008

Nguyen DX, Bos PD, Massagué J. Metastasis: from dissemination to organ-specific colonization. Nat Rev Cancer. **2009** Apr;9(4):274-84. DOI: 10.1038/nrc2622

Nguyen A, Yoshida M, Goodarzi H, Tavazoie SF. Highly variable cancer subpopulations that exhibit enhanced transcriptome variability and metastatic fitness. Nat Commun. **2016** May 3;7:11246. DOI: 10.1038/ncomms11246

Ni C and Huang J. Dynamic regulation of cancer stem cells and clinical challenges. Clin Transl Oncol. **2013** Apr;15(4):253-8. DOI: 10.1007/s12094-012-0927-7

Nielsen TO, Parker JS, Leung S, Voduc D, Ebbert M, Vickery T, Davies SR, Snider J, Stijleman IJ, Reed J, Cheang MC, Mardis ER, Perou CM, Bernard PS, Ellis MJ. A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer. Clin Cancer Res. **2010** Nov 1;16(21):5222-32. DOI: 10.1158/1078-0432.CCR-10-1282

Nieswandt B, Hafner M, Echtenacher B, Männel DN. Lysis of tumor cells by natural killer cells in mice is impeded by platelets. Cancer Res. **1999** Mar 15;59(6):1295-300. PMID: 10096562

Nieto MA, Huang RY, Jackson RA, Thiery JP. EMT: 2016. Cell. **2016** Jun 30;166(1):21-45. DOI: 10.1016/j.cell.2016.06.028

Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, Martincorena I, Alexandrov LB, Martin S, Wedge DC, Van Loo P, Ju YS, Smid M, Brinkman AB, Morganella S, Aure MR, Lingjærde OC, Langerød A, Ringnér M, Ahn SM, Boyault S, Brock JE, Broeks A, Butler A, Desmedt C, Dirix L, Dronov S, Fatima A, Foekens JA, Gerstung M, Hooijer GK, Jang SJ, Jones DR, Kim HY, King TA, Krishnamurthy S, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature. **2016** May 2;534(7605):47-54. DOI: 10.1038/nature17676

Norton L and Massagué J. Is cancer a disease of self-seeding? Nat Med. **2006** Aug;12(8):875-8. DOI: 10.1038/nm0806-875

Norton KA, Popel AS, Pandey NB. Heterogeneity of chemokine cell-surface receptor expression in triple-negative breast cancer. Am J Cancer Res. **2015** Mar 15;5(4):1295-307. PMCID: PMC4473311

Nowell PC. The clonal evolution of tumor cell populations. Science. **1976** Oct 1;194(4260):23-8. PMID: 959840

Ntouroupi TG, Ashraf SQ, McGregor SB, Turney BW, Seppo A, Kim Y, Wang X, Kilpatrick MW, Tsipouras P, Tafas T, Bodmer WF. Detection of circulating tumour cells in peripheral blood with an automated scanning fluorescence microscope. Br J Cancer. **2008** Sep 2;99(5):789-95. DOI: 10.1038/sj.bjc.6604545

Oakes SR, Gallego-Ortega D, Ormandy CJ. The mammary cellular hierarchy and breast cancer. Cell Mol Life Sci. **2014** Nov;71(22):4301-24. DOI: 10.1007/s00018-014-1674-4

Oakes SR, Hilton HN, Ormandy CJ. The alveolar switch: coordinating the proliferative cues and cell fate decisions that drive the formation of lobuloalveoli from ductal epithelium. Breast Cancer Res. **2006**;8(2):207. DOI: 10.1186/bcr1411

Ocaña OH, Córcoles R, Fabra A, Moreno-Bueno G, Acloque H, Vega S, Barrallo-Gimeno A, Cano A, Nieto. Metastatic colonization requires the repression of the epithelial-mesenchymal transition inducer Prrx1. Cancer Cell. **2012** Dec 11;22(6):709-24. DOI: 10.1016/j.ccr.2012.10.012

Ocaña A, Diez-Gónzález L, Adrover E, Fernández-Aramburo A, Pandiella A, Amir E. Tumor-infiltrating lymphocytes in breast cancer: ready for prime time? J Clin Oncol. **2015** Apr 10;33(11):1298-9. DOI: 10.1200/JCO.2014.59.7286

Oehme I, Linke JP, Böck BC, Milde T, Lodrini M, Hartenstein B, Wiegand I, Eckert C, Roth W, Kool M, Kaden S, Gröne HJ, Schulte JH, Lindner S, Hamacher-Brady A, Brady NR, Deubzer HE, Witt O. Histone deacetylase 10 promotes autophagy-mediated cell survival. Proc Natl Acad Sci U S A. **2013** Jul 9;110(28):E2592-601. DOI: 10.1073/pnas.1300113110

Ogata H, Inoue N, Podolsky DK. Identification of a goblet cell-specific enhancer element in the rat intestinal trefoil factor gene promoter bound by a goblet cell nuclear protein. J Biol Chem. **1998** Jan 30;273(5):3060-7. PMID: 9446622

Oskarsson T, Batlle E, Massagué J. Metastatic stem cells: sources, niches, and vital pathways. Cell Stem Cell. **2014** Mar 6;14(3):306-21. DOI: 10.1016/j.stem.2014.02.002

Oyama T, Maluf H, Koerner F. Atypical cystic lobules: an early stage in the formation of low-grade ductal carcinoma in situ. Virchows Arch. **1999** Oct;435(4):413-21. PMID: 10526005

Padua D, Zhang XH, Wang Q, Nadal C, Gerald WL, Gomis RR, Massagué J. TGFbeta primes breast tumors for lung metastasis seeding through angiopoietin-like 4. Cell. **2008** Apr 4;133(1):66-77. DOI: 10.1016/j.cell.2008.01.046

Padua D and Massagué J. Roles of TGFbeta in metastasis. Cell Res. **2009** Jan;19(1):89-102. DOI: 10.1038/cr.2008.316.

Page DL and Rogers LW. Combined histologic and cytologic criteria for the diagnosis of mammary atypical ductal hyperplasia. Hum Pathol. **1992** Oct;23(10):1095-7. PMID: 1328030

Paget S. The distribution of secondary growths in cancer of the breast. Lancet. **1889** March 23;133(3421):571-3 (originally published as volume 1, issue 3421). DOI: 10.1016/S0140-6736(00)49915-0

Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, Park T, Hiller W, Fisher ER, Wickerham DL, Bryant J, Wolmark N. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med. **2004** Dec 30;351(27):2817-26. DOI: 10.1056/NEJMoa041588

Paik S, Tang G, Shak S, Kim C, Baker J, Kim W, Cronin M, Baehner FL, Watson D, Bryant J, Costantino JP, Geyer CE Jr, Wickerham DL, Wolmark N. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. J Clin Oncol. **2006** Aug 10;24(23):3726-34. DOI: 10.1200/JCO.2005.04.7985

Pandey V, Wu ZS, Zhang M, Li R, Zhang J, Zhu T, Lobie PE. Trefoil factor 3 promotes metastatic seeding and predicts poor survival outcome of patients with mammary carcinoma. Breast Cancer Res. **2014** Sep 30;16(5):429. DOI: 10.1186/s13058-014-0429-3

Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z, Quackenbush JF, Stijleman IJ, Palazzo J, Marron JS, Nobel AB, Mardis E, Nielsen TO, Ellis MJ, Perou CM, Bernard PS. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol. **2009** Mar 10;27(8):1160-7. DOI: 10.1200/JCO.2008.18.1370

Patrawala L, Calhoun T, Schneider-Broussard R, Zhou J, Claypool K, Tang DG. Side population is enriched in tumorigenic, stem-like cancer cells, whereas ABCG2+ and ABCG2- cancer cells are similarly tumorigenic. Cancer Res. **2005** Jul 15;65(14):6207-19. DOI: 10.1158/0008-5472.CAN-05-0592

Pattabiraman DR and Weinberg RA. Tackling the cancer stem cells - what challenges do they pose? Nat Rev Drug Discov. **2014** Jul;13(7):497-512. DOI: 10.1038/nrd4253

Pawelek JM and Chakraborty AK. The cancer cell--leukocyte fusion theory of metastasis. Adv Cancer Res. **2008**;101:397-444. DOI: 10.1016/S0065-230X(08)00410-7

Peeters DJ, De Laere B, Van den Eynden GG, Van Laere SJ, Rothé F, Ignatiadis M, Sieuwerts AM, Lambrechts D, Rutten A, van Dam PA, Pauwels P, Peeters M, Vermeulen PB, Dirix LY. Semiautomated isolation and molecular characterisation of single or highly purified tumour cells from CellSearch enriched blood samples using dielectrophoretic cell sorting. Br J Cancer. **2013** Apr 2;108(6):1358-67. DOI: 10.1038/bjc.2013.92

Perera O, Evans A, Pertziger M, MacDonald C, Chen H, Liu DX, Lobie PE, Perry JK. Trefoil factor 3 (TFF3) enhances the oncogenic characteristics of prostate carcinoma cells and reduces sensitivity to ionising radiation. Cancer Lett. **2015** May 28;361(1):104-11. DOI: 10.1016/j.canlet.2015.02.051

Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lønning PE, Børresen-Dale AL, Brown PO, Botstein D. Molecular portraits of human breast tumours. Nature. **2000** Aug 17;406(6797):747-52. DOI:10.1038/35021093

Perry JK, Kannan N, Grandison PM, Mitchell MD, Lobie PE. Are trefoil factors oncogenic? Trends Endocrinol Metab. **2008** Mar;19(2):74-81. DOI: 10.1016/j.tem.2007.10.003

Pestrin M, Bessi S, Galardi F, Truglia M, Biggeri A, Biagioni C, Cappadona S, Biganzoli L, Giannini A, Di Leo A. Correlation of HER2 status between primary tumors and corresponding circulating tumor cells in advanced breast cancer patients. Breast Cancer Res Treat. **2009** Dec;118(3):523-30. DOI: 10.1007/s10549-009-0461-7

Pestrin M, Bessi S, Puglisi F, Minisini AM, Masci G, Battelli N, Ravaioli A, Gianni L, Di Marsico R, Tondini C, Gori S, Coombes CR, Stebbing J, Biganzoli L, Buyse M, Di Leo A. Final results of a multicenter phase II clinical trial evaluating the activity of single-agent lapatinib in patients with HER2-negative metastatic breast cancer and HER2-positive circulating tumor cells. A proof-of-concept study. Breast Cancer Res Treat. **2012** Jul;134(1):283-9. DOI: 10.1007/s10549-012-2045-1

Pestrin M, Salvianti F, Galardi F, De Luca F, Turner N, Malorni L, Pazzagli M, Di Leo A, Pinzani P. Heterogeneity of PIK3CA mutational status at the single cell level in circulating tumor cells from metastatic breast cancer patients. Mol Oncol. **2015** Apr;9(4):749-57. DOI: 10.1016/j.molonc.2014.12.001

Peto R, Davies C, Godwin J, Gray R, Pan HC, Clarke M, Cutter D, Darby S, McGale P, Taylor C, Wang YC, Bergh J, Di Leo A, Albain K, Swain S, Piccart M, Pritchard K, et al. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Comparisons between different

polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100,000 women in 123 randomised trials. Lancet. **2012** Feb 4;379(9814):432-44. DOI: 10.1016/S0140-6736(11)61625-5

Phillips TM, McBride WH, Pajonk F. The response of CD24(-/low)/CD44+ breast cancer-initiating cells to radiation. J Natl Cancer Inst. **2006** Dec 20;98(24):1777-85. DOI: 10.1093/jnci/djj495

Piccart M, Rutgers E, van't Veer L, et al: Primary analysis of the EORTC 10041/BIG 3-04 MINDACT study: A prospective, randomized study evaluating the clinical utility of the 70-gene signature (MammaPrint) combined with common clinical-pathological criteria for selection of patients for adjuvant chemotherapy in breast cancer with 0 to 3 positive nodes. **2016** American Association of Cancer Research Annual Meeting. Abstract CT039.

Pierga JY, Bidard FC, Mathiot C, Brain E, Delaloge S, Giachetti S, de Cremoux P, Salmon R, Vincent-Salomon A, Marty M. Circulating tumor cell detection predicts early metastatic relapse after neoadjuvant chemotherapy in large operable and locally advanced breast cancer in a phase II randomized trial. Clin Cancer Res. **2008** Nov 1;14(21):7004-10. DOI: 10.1158/1078-0432.CCR-08-0030

Podolsky DK, Lynch-Devaney K, Stow JL, Oates P, Murgue B, De-Beaumont M, Sands BE, Mahida YR. Identification of human intestinal trefoil factor. Goblet cell-specific expression of a peptide targeted for apical secretion. J Biol Chem. **1993** Jun 5;268(16):12230. PMID: 8505343

Podsypanina K, Du YC, Jechlinger M, Beverly LJ, Hambardzumyan D, Varmus H. Seeding and propagation of untransformed mouse mammary cells in the lung. Science. **2008** Sep 26;321(5897):1841-4. DOI: 10.1126/science.1161621

Polyak K. On the birth of breast cancer. Biochim Biophys Acta. **2001** Nov 30;1552(1):1-13. DOI:10.1016/S0304-419X(01)00029-4

Polyak K. Heterogeneity in breast cancer. J Clin Invest. **2011** Oct;121(10):3786-8. DOI: 10.1172/JCI60534

Ponder BA. Cancer genetics. Nature. **2001** May 17;411(6835):336-41. DOI: 10.1038/35077207

Ponti D, Costa A, Zaffaroni N, Pratesi G, Petrangolini G, Coradini D, Pilotti S, Pierotti MA, Daidone MG. Isolation and in vitro propagation of tumorigenic breast cancer cells with stem/progenitor cell properties. Cancer Res. **2005** Jul 1;65(13):5506-11. DOI: 10.1158/0008-5472.CAN-05-0626

Poste G and Fidler IJ. Nature. **1980** Jan 10;283(5743):139-46. The pathogenesis of cancer metastasis. PMID: 6985715

Poulsom R and Wright NA. Trefoil peptides: a newly recognized family of epithelial mucin-associated molecules. Am J Physiol. **1993** Aug;265(2 Pt 1):G205-13.

Poulsom R, Begos DE, Modlin IM. Molecular aspects of restitution: functions of trefoil peptides. Yale J Biol Med. **1996** Mar-Apr;69(2):137-46. PMCID: PMC2588989

Poulsom R, Hanby AM, Lalani EN, Hauser F, Hoffmann W, Stamp GW. Intestinal trefoil factor (TFF3) and pS2 (TFF1), but not spasmolytic polypeptide (TFF2) mRNAs are co-expressed in normal, hyperplastic, and neoplastic human breast epithelium. Pathol. **1997** Sep;183(1):30-8. DOI: 10.1002/(SICI)1096-9896(199709)183:1<30::AID-PATH1085>3.0.CO;2-K

Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, He X, Perou CM. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res. **2010**;12(5):R68. DOI: 10.1186/bcr2635

Price JE, Polyzos A, Zhang RD, Daniels LM. Tumorigenicity and metastasis of human breast carcinoma cell lines in nude mice. Cancer Res. **1990** Feb 1;50(3):717-21. PMID: 2297709

Quail DF and Joyce JA. Microenvironmental regulation of tumor progression and metastasis. Nat Med. **2013** Nov;19(11):1423-37. DOI: 10.1038/nm.3394

Rack B, Schindlbeck C, Jückstock J, Andergassen U, Hepp P, Zwingers T, Friedl TW, Lorenz R, Tesch H, Fasching PA, Fehm T, Schneeweiss A, Lichtenegger W, Beckmann MW, Friese K, Pantel K, Janni W; SUCCESS Study Group. Circulating tumor cells predict survival in early average-to-high risk breast cancer patients. J Natl Cancer Inst. **2014** May 15;106(5). pii: dju066. DOI: 10.1093/jnci/dju066

Rahman M, Davis SR, Pumphrey JG, Bao J, Nau MM, Meltzer PS, Lipkowitz S. TRAIL induces apoptosis in triple-negative breast cancer cells with a mesenchymal phenotype. Breast Cancer Res Treat. **2009** Jan;113(2):217-30. DOI: 10.1007/s10549-008-9924-5

Ralhan R, Kaur J, Kreienberg R, Wiesmüller L. Links between DNA double strand break repair and breast cancer: accumulating evidence from both familial and nonfamilial cases. Cancer Lett. **2007** Apr 8;248(1):1-17. DOI: 10.1016/j.canlet.2006.06.004

Ramirez JM, Fehm T, Orsini M, Cayrefourcq L, Maudelonde T, Pantel K, Alix-Panabières C. Prognostic relevance of viable circulating tumor cells detected by EPISPOT in metastatic breast cancer patients. Clin Chem. **2014** Jan;60(1):214-21. DOI: 10.1373/clinchem.2013.215079

Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. Nat Protoc. **2013** Nov;8(11):2281-308. DOI: 10.1038/nprot.2013.143

Récamier JCA, Recherches sur le traitement du cancer par la compression méthodique simple ou combinée, et sur l'histoire général de la même maladie. **1829**. Paris, Gabon

Reeves EP, Ali T, Leonard P, Hearty S, O'Kennedy R, May FE, Westley BR, Josenhans C, Rust M, Suerbaum S, Smith A, Drumm B, Clyne M. Helicobacter pylori lipopolysaccharide interacts with TFF1 in a pH-dependent manner. Gastroenterology. **2008** Dec;135(6):2043-54, 2054.e1-2. DOI: 10.1053/j.gastro.2008.08.049

Regalo G, Wright NA, Machado JC. Trefoil factors: from ulceration to neoplasia. Cell Mol Life Sci. **2005** Dec;62(24):2910-5. DOI: 10.1007/s00018-005-5478-4

Reiske HR, Kao SC, Cary LA, Guan JL, Lai JF, Chen HC. Requirement of phosphatidylinositol 3-kinase in focal adhesion kinase-promoted cell migration. J Biol Chem. **1999** Apr 30;274(18):12361-6. PMID: 10212207

Resetkova E, Reis-Filho JS, Jain RK, Mehta R, Thorat MA, Nakshatri H, Badve S. Prognostic impact of ALDH1 in breast cancer: a story of stem cells and tumor microenvironment. Breast Cancer Res Treat. **2010** Aug;123(1):97-108. DOI: 10.1007/s10549-009-0619-3

Richert MM, Schwertfeger KL, Ryder JW, Anderson SM. An atlas of mouse mammary gland development. J Mammary Gland Biol Neoplasia. **2000** Apr;5(2):227-41. DOI: 10.1023/A:1026499523505

Riethdorf S, Fritsche H, Müller V, Rau T, Schindlbeck C, Rack B, Janni W, Coith C, Beck K, Jänicke F, Jackson S, Gornet T, Cristofanilli M, Pantel K. Detection of circulating tumor cells in peripheral blood of patients with metastatic breast cancer: a validation study of the CellSearch system. Clin Cancer Res. **2007** Feb 1;13(3):920-8. DOI: 10.1158/1078-0432.CCR-06-1695

Rios AC, Fu NY, Lindeman GJ, Visvader JE. In situ identification of bipotent stem cells in the mammary gland. Nature. **2014** Feb 20;506(7488):322-7. DOI: 10.1038/nature12948

Rivat C, Rodrigues S, Bruyneel E, Piétu G, Robert A, Redeuilh G, Bracke M, Gespach C, Attoub S. Cancer Res. Implication of STAT3 signaling in human colonic cancer cells during intestinal trefoil factor 3 (TFF3) -- and vascular endothelial growth factor-mediated cellular invasion and tumor growth. **2005** Jan 1;65(1):195-202. PMID: 15665295

Robinson DR, Wu YM, Vats P, Su F, Lonigro RJ, Cao X, Kalyana-Sundaram S, Wang R, Ning Y, Hodges L, Gursky A, Siddiqui J, Tomlins SA, Roychowdhury S, Pienta KJ, Kim SY, Roberts JS, Rae JM, Van Poznak CH, Hayes DF, Chugh R, Kunju LP, Talpaz M, Schott AF, Chinnaiyan AMNat Genet. Activating ESR1 mutations in hormone-resistant metastatic breast cancer. **2013** Dec;45(12):1446-51. DOI: 10.1038/ng.2823

Rodrigues S, Nguyen QD, Faivre S, Bruyneel E, Thim L, Westley B, May F, Flatau G, Mareel M, Gespach C, Emami S. Activation of cellular invasion by trefoil peptides and src is mediated by cyclooxygenase- and thromboxane A2 receptor-dependent signaling pathways. FASEB J. **2001** Jul;15(9):1517-28. DOI: 10.1096/fj.00-0802com

Rodrigues S, Van Aken E, Van Bocxlaer S, Attoub S, Nguyen QD, Bruyneel E, Westley BR, May FE, Thim L, Mareel M, Gespach C, Emami S. Trefoil peptides as proangiogenic factors in vivo and in vitro: implication of cyclooxygenase-2 and EGF receptor signaling. FASEB J. **2003a** Jan;17(1):7-16. DOI: 10.1096/fj.02-0201com

Rodrigues S, Attoub S, Nguyen QD, Bruyneel E, Rodrigue CM, Westley BR, May FE, Thim L, Mareel M, Emami S, Gespach C. Selective abrogation of the proinvasive activity of the trefoil peptides pS2 and spasmolytic polypeptide by disruption of the EGF receptor signaling pathways in kidney and colonic cancer cells. Oncogene. **2003b** Jul 17;22(29):4488-97. DOI: 10.1038/sj.onc.1206685

Rodrigues P, Hering FO, Meller A. Adjuvant Effect of IV Clodronate on the Delay of Bone Metastasis in High-Risk Prostate Cancer Patients: A Prospective Study. Cancer Res Treat. **2011** Dec;43(4):231-5. DOI: 10.4143/crt.2011.43.4.231

Roodman GD. Mechanisms of bone metastasis. N Engl J Med. **2004** Apr 15;350(16):1655-64. DOI: 10.1056/NEJMra030831

Roschke AV and Rozenblum E. Multi-layered cancer chromosomal instability phenotype. Front Oncol. **2013** Dec 11;3:302. DOI: 10.3389/fonc.2013.00302

Rossi A, Kontarakis Z, Gerri C, Nolte H, Hölper S, Krüger M, Stainier DY. Genetic compensation induced by deleterious mutations but not gene knockdowns. Nature. **2015** Aug 13;524(7564):230-3. DOI: 10.1038/nature14580

Sakakura T. Embryogenesis. Development, Regulation and Function. **1987**. In: Neville M.C., Daniel C.W., editors. New York: Plenum; pp 37–65

Saloustros E, Perraki M, Apostolaki S, Kallergi G, Xyrafas A, Kalbakis K, Agelaki S, Kalykaki A, Georgoulias V, Mavroudis D. Cytokeratin-19 mRNA-positive circulating tumor cells during follow-up of patients with operable breast cancer: prognostic relevance for late relapse. Breast Cancer Res. **2011** Jun 10;13(3):R60. DOI: 10.1186/bcr2897

Santa-Maria CA and Gradishar WJ. Changing Treatment Paradigms in Metastatic Breast Cancer: Lessons Learned. JAMA Oncol. **2015** Jul;1(4):528-34; quiz 549. DOI: 10.1001/jamaoncol.2015.1198

Santilli G, Binda M, Zaffaroni N, Daidone MG. Breast cancer-initiating cells: insights into novel treatment strategies. Cancers (Basel). **2011** Mar 16;3(1):1405-25. DOI: 10.3390/cancers3011405

Sasportas LS and Gambhir SS. Imaging circulating tumor cells in freely moving awake small animals using a miniaturized intravital microscope. PLoS One. **2014** Jan 31;9(1):e86759. DOI: 10.1371/journal.pone.0086759

Saxena M and Christofori G. Rebuilding cancer metastasis in the mouse. Mol Oncol. **2013** Apr;7(2):283-96. DOI: 10.1016/j.molonc.2013.02.009

Schardt JA, Meyer M, Hartmann CH, Schubert F, Schmidt-Kittler O, Fuhrmann C, Polzer B, Petronio M, Eils R, Klein CA. Genomic analysis of single cytokeratin-positive cells from bone marrow reveals early mutational events in breast cancer. Cancer Cell. **2005** Sep;8(3):227-39. DOI: 10.1016/j.ccr.2005.08.003

Schott AF, Hayes DF. Defining the benefits of neoadjuvant chemotherapy for breast cancer. J Clin Oncol. **2012** May 20;30(15):1747-9. DOI: 10.1200/JCO.2011.41.3161

Schug ZT, Peck B, Jones DT, Zhang Q, Grosskurth S, Alam IS, Goodwin LM, Smethurst E, Mason S, Blyth K, McGarry L, James D, Shanks E, Kalna G, Saunders RE, Jiang M, Howell M, Lassailly F, Thin MZ, Spencer-Dene B, Stamp G, van den Broek NJ, Mackay G, Bulusu V, Kamphorst JJ, Tardito S, Strachan D, Harris AL, Aboagye EO, Critchlow SE, Wakelam MJ, Schulze A, Gottlieb E. Acetyl-CoA synthetase 2 promotes acetate utilization and maintains

cancer cell growth under metabolic stress. Cancer Cell. **2015** Jan 12;27(1):57-71. DOI: 10.1016/j.ccell.2014.12.002

Senkus E, Kyriakides S, Ohno S, Penault-Llorca F, Poortmans P, Rutgers E, Zackrisson S, Cardoso F; ESMO Guidelines Committee. Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. **2015** Sep;26 Suppl 5:v8-30. DOI: 10.1093/annonc/mdv298

Shackleton M, Vaillant F, Simpson KJ, Stingl J, Smyth GK, Asselin-Labat ML, Wu L, Lindeman GJ, Visvader JE. Generation of a functional mammary gland from a single stem cell. Nature. **2006** Jan 5;439(7072):84-8. DOI: 10.1038/nature04372

Shah SP, Morin RD, Khattra J, Prentice L, Pugh T, Burleigh A, Delaney A, Gelmon K, Guliany R, Senz J, Steidl C, Holt RA, Jones S, Sun M, Leung G, Moore R, Severson T, Taylor GA, Teschendorff AE, Tse K, Turashvili G, Varhol R, Warren RL, Watson P, Zhao Y, Caldas C, Huntsman D, Hirst M, Marra MA, Aparicio S. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. Nature. **2009** Oct 8;461(7265):809-13. DOI: 10.1038/nature08489

Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, Turashvili G, Ding J, Tse K, Haffari G, Bashashati A, Prentice LM, Khattra J, Burleigh A, Yap D, Bernard V, McPherson A, Shumansky K, Crisan A, Giuliany R, Heravi-Moussavi A, Rosner J, Lai D, Birol I, Varhol R, Tam A, Dhalla N, Zeng T, Ma K, Chan SK, Griffith M, Moradian A, Cheng SW, Morin GB, Watson P, Gelmon K, Chia S, Chin SF, Curtis C, Rueda OM, Pharoah PD, Damaraju S, Mackey J, Hoon K, Harkins T, Tadigotla V, Sigaroudinia M, Gascard P, Tlsty T, Costello JF, Meyer IM, Eaves CJ, Wasserman WW, Jones S, Huntsman D, Hirst M, Caldas C, Marra MA, Aparicio S. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature. **2012** Apr 4;486(7403):395-9. DOI: 10.1038/nature10933

Shaw LM. Tumor cell invasion assays. Methods Mol Biol. **2005**;294:97-105. PMID: 15576908

Sherman A, Dawson A, Mather C, Gilhooley H, Li Y, Mitchell R, Finnegan D, Sang H. Transposition of the Drosophila element mariner into the chicken germ line. Nat Biotechnol. **1998** Nov;16(11):1050-3. DOI: 10.1038/3497

Shou J, Massarweh S, Osborne CK, Wakeling AE, Ali S, Weiss H, Schiff R. Mechanisms of tamoxifen resistance: increased estrogen receptor-HER2/neu cross-talk in ER/HER2-positive breast cancer. J Natl Cancer Inst. **2004** Jun 16;96(12):926-35. DOI: 10.1093/jnci/djh166

Simões BM, O'Brien CS, Eyre R, Silva A, Yu L, Sarmiento-Castro A, Alférez DG, Spence K, Santiago-Gómez A, Chemi F, Acar A, Gandhi A, Howell A, Brennan K, Rydén L, Catalano S, Andó S, Gee J, Ucar A, Sims AH, Marangoni E, Farnie G, Landberg G, Howell SJ, Clarke RB. Anti-estrogen Resistance in Human Breast Tumors Is Driven by JAG1-NOTCH4-Dependent Cancer Stem Cell Activity. Cell Rep. **2015** Sep 29;12(12):1968-77. DOI: 10.1016/j.celrep.2015.08.050

Singh JK, Farnie G, Bundred NJ, Simões BM, Shergill A, Landberg G, Howell SJ, Clarke RB. Targeting CXCR1/2 significantly reduces breast cancer stem cell activity and increases the efficacy of inhibiting HER2 via HER2-dependent and -independent mechanisms. Clin Cancer Res. **2013** Feb 1;19(3):643-56. DOI: 10.1158/1078-0432.CCR-12-1063

Singletary SE. Rating the risk factors for breast cancer. Ann Surg. **2003** Apr;237(4):474-82. DOI: 10.1097/01.SLA.0000059969.64262.87

Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science. **1987** Jan 9;235(4785):177-82. PMID: 3798106

Slamon DJ, Godolphin W, Jones LA, Holt JA, Wong SG, Keith DE, Levin WJ, Stuart SG, Udove J, Ullrich A, et al. Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. Science. **1989** May 12;244(4905):707-12. PMID: 2470152

Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, Fleming T, Eiermann W, Wolter J, Pegram M, Baselga J, Norton L. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. N Engl J Med. **2001** Mar 15;344(11):783-92. DOI: 10.1056/NEJM200103153441101

Sleeman JP and Thiery JP. SnapShot: The epithelial-mesenchymal transition. **Cell**. 2011 Apr 1;145(1):162.e1. DOI: 10.1016/j.cell.2011.03.029

Smerage JB, Barlow WE, Hortobagyi GN, Winer EP, Leyland-Jones B, Srkalovic G, Tejwani S, Schott AF, O'Rourke MA, Lew DL, Doyle GV, Gralow JR, Livingston RB, Hayes DF. Circulating tumor cells and response to chemotherapy in metastatic breast cancer: SWOG S0500. J Clin Oncol. **2014** Nov 1;32(31):3483-9. DOI: 10.1200/JCO.2014.56.2561

Smid M, Wang Y, Klijn JG, Sieuwerts AM, Zhang Y, Atkins D, Martens JW, Foekens JA. Genes associated with breast cancer metastatic to bone. J Clin Oncol. **2006** May 20;24(15):2261-7. DOI: 10.1200/JCO.2005.03.8802

Smirnov DA, Zweitzig DR, Foulk BW, Miller MC, Doyle GV, Pienta KJ, Meropol NJ, Weiner LM, Cohen SJ, Moreno JG, Connelly MC, Terstappen LW, O'Hara SM. Global gene expression profiling of circulating tumor cells. Cancer Res. **2005** Jun 15;65(12):4993-7. DOI: 10.1158/0008-5472.CAN-04-4330

Smith GH. Label-retaining epithelial cells in mouse mammary gland divide asymmetrically and retain their template DNA strands. Development. **2005** Feb;132(4):681-7. DOI: 10.1242/dev.01609

Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol. **2004**; 3:Article3. DOI: 10.2202/1544-6115.1027

Sommer S and Fuqua SA. Estrogen receptor and breast cancer. Semin Cancer Biol. **2001** Oct;11(5):339-52.

Song C, Zhu S, Wu C, Kang J. Histone deacetylase (HDAC) 10 suppresses cervical cancer metastasis through inhibition of matrix metalloproteinase (MMP) 2 and 9 expression. J Biol Chem. **2013** Sep 27;288(39):28021-33. DOI: 10.1074/jbc.M113.498758

Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, Lønning PE, Brown PO, Børresen-Dale AL, Botstein D.

Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A. **2003** Jul 8;100(14):8418-23. DOI: 10.1073/pnas.0932692100

Sosa MS, Avivar-Valderas A, Bragado P, Wen HC, Aguirre-Ghiso JA. ERK1/2 and p38α/β signaling in tumor cell quiescence: opportunities to control dormant residual disease. Clin Cancer Res. **2011** Sep 15;17(18):5850-7. DOI: 10.1158/1078-0432.CCR-10-2574

Sosa MS, Bragado P, Aguirre-Ghiso JA. Mechanisms of disseminated cancer cell dormancy: an awakening field. Nat Rev Cancer. **2014** Sep;14(9):611-22. DOI: 10.1038/nrc3793

Sparano JA and Paik S. Development of the 21-gene assay and its application in clinical practice and clinical trials. J Clin Oncol. **2008** Feb 10;26(5):721-8. DOI: 10.1200/JCO.2007.15.1068

Sparano JA, Gray RJ, Makower DF, Pritchard KI, Albain KS, Hayes DF, Geyer CE Jr, Dees EC, Perez EA, Olson JA Jr, Zujewski J, Lively T, Badve SS, Saphner TJ, Wagner LI, Whelan TJ, Ellis MJ, Paik S, Wood WC, Ravdin P, Keane MM, Gomez Moreno HL, Reddy PS, Goggins TF, Mayer IA, Brufsky AM, Toppmeyer DL, Kaklamani VG, Atkins JN, Berenberg JL, Sledge GW. Prospective Validation of a 21-Gene Expression Assay in Breast Cancer. N Engl J Med. **2015** Nov 19;373(21):2005-14. DOI: 10.1056/NEJMoa1510764

Spike BT, Engle DD, Lin JC, Cheung SK, La J, Wahl GM. A mammary stem cell population identified and characterized in late embryogenesis reveals similarities to human breast cancer. Cell Stem Cell. **2012** Feb 3;10(2):183-97. DOI: 10.1016/j.stem.2011.12.018

Stacer AC, Fenner J, Cavnar SP, Xiao A, Zhao S, Chang SL, Salomonnson A, Luker KE, Luker GD. Endothelial CXCR7 regulates breast cancer metastasis. Oncogene. **2016** Mar 31;35(13):1716-24. DOI: 10.1038/onc.2015.236

Steeg PS. Tumor metastasis: mechanistic insights and clinical challenges. Nat Med. **2006** Aug;12(8):895-904. DOI: 10.1038/nm1469

Stingl J, Eirew P, Ricketson I, Shackleton M, Vaillant F, Choi D, Li HI, Eaves CJ. Purification and unique properties of mammary epithelial stem cells. Nature. **2006** Feb 23;439(7079):993-7. DOI: 10.1038/nature04496

Stoker M and Perryman M. An epithelial scatter factor released by embryo fibroblasts. J Cell Sci. **1985** Aug;77:209-23. PMID: 3841349

Stott SL, Hsu CH, Tsukrov DI, Yu M, Miyamoto DT, Waltman BA, Rothenberg SM, Shah AM, Smas ME, Korir GK, Floyd FP Jr, Gilman AJ, Lord JB, Winokur D, Springer S, Irimia D, Nagrath S, Sequist LV, Lee RJ, Isselbacher KJ, Maheswaran S, Haber DA, Toner M. Isolation of circulating tumor cells using a microvortex-generating herringbone-chip. Proc Natl Acad Sci U S A. **2010** Oct 26;107(43):18392-7. DOI: 10.1073/pnas.1012539107

Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature. **2009** Apr 9;458(7239):719-24. DOI: 10.1038/nature07943

Stuckey A, Febbraro T, Laprise J, Wilbur JS, Lopes V, Robison K. Adherence Patterns to National Comprehensive Cancer Network Guidelines for Referral of Women With Breast

Cancer to Genetics Professionals. Am J Clin Oncol. **2016** Aug;39(4):363-7. DOI: 10.1097/COC.0000000000000073

Stute P, Wood CE, Kaplan JR, Cline JM. Cyclic changes in the mammary gland of cynomolgus macaques. Fertil Steril. **2004** Oct;82 Suppl 3:1160-70. DOI: 10.1016/j.fertnstert.2004.04.035

Stylianou S, Clarke RB, Brennan K. Aberrant activation of notch signaling in human breast cancer. Cancer Res. **2006** Feb 1;66(3):1517-25. DOI: 10.1158/0008-5472.CAN-05-3054

Su S, Liu Q, Chen J, Chen J, Chen F, He C, Huang D, Wu W, Lin L, Huang W, Zhang J, Cui X, Zheng F, Li H, Yao H, Su F, Song E. A positive feedback loop between mesenchymal-like cancer cells and macrophages is essential to breast cancer metastasis. Cancer Cell. **2014** May 12;25(5):605-20. DOI: 10.1016/j.ccr.2014.03.021

Su X, Malouf GG, Chen Y, Zhang J, Yao H, Valero V, Weinstein JN, Spano JP, Meric-Bernstam F, Khayat D, Esteva FJ. Comprehensive analysis of long non-coding RNAs in human breast cancer clinical subtypes. Oncotarget. **2014** Oct 30;5(20):9864-76. DOI: 10.18632/oncotarget.2454

Suemori S, Lynch-Devaney K, Podolsky DK. Identification and characterization of rat intestinal trefoil factor: tissue- and cell-specific member of the trefoil protein family. Proc Natl Acad Sci U S A. **1991** Dec 15;88(24):11017-21. PMCID: PMC53064

Sun Y, Boyd K, Xu W, Ma J, Jackson CW, Fu A, Shillingford JM, Robinson GW, Hennighausen L, Hitzler JK, Ma Z, Morris SW. Acute myeloid leukemia-associated Mkl1 (Mrtf-a) is a key regulator of mammary gland function. Mol Cell Biol. **2006** Aug;26(15):5809-26. DOI: 10.1128/MCB.00024-06

Swain SM, Baselga J, Kim SB, Ro J, Semiglazov V, Campone M, Ciruelos E, Ferrero JM, Schneeweiss A, Heeson S, Clark E, Ross G, Benyunes MC, Cortés J; CLEOPATRA Study Group. Pertuzumab, trastuzumab, and docetaxel in HER2-positive metastatic breast cancer. N Engl J Med. **2015** Feb 19;372(8):724-34. DOI: 10.1056/NEJMoa1413513

Tabassum DP and Polyak K. Tumorigenesis: it takes a village. Nat Rev Cancer. **2015** Aug;15(8):473-83. DOI: 10.1038/nrc3971

Tabernero J, Rojo F, Calvo E, Burris H, Judson I, Hazell K, Martinelli E, Ramon y Cajal S, Jones S, Vidal L, Shand N, Macarulla T, Ramos FJ, Dimitrijevic S, Zoellner U, Tang P, Stumm M, Lane HA, Lebwohl D, Baselga J. Dose- and schedule-dependent inhibition of the mammalian target of rapamycin pathway with everolimus: a phase I tumor pharmacodynamic study in patients with advanced solid tumors. J Clin Oncol. **2008** Apr 1;26(10):1603-10. DOI: 10.1200/JCO.2007.14.5482

Talmadge JE and Fidler IJ. AACR centennial series: the biology of cancer metastasis: historical perspective. Cancer Res. **2010** Jul 15;70(14):5649-69. DOI: 10.1158/0008-5472.CAN-10-1040

Taupin DR, Kinoshita K, Podolsky DK. Intestinal trefoil factor confers colonic epithelial resistance to apoptosis. Proc Natl Acad Sci U S A. **2000** Jan 18;97(2):799-804. PMCID: PMC15411

Taupin D and Podolsky DK. Trefoil factors: initiators of mucosal healing. Nat Rev Mol Cell Biol. **2003** Sep;4(9):721-32. DOI: 10.1038/nrm1203

Teschendorff AE, Miremadi A, Pinder SE, Ellis IO, Caldas C. An immune response gene expression module identifies a good prognosis subtype in estrogen receptor negative breast cancer. Genome Biol. **2007**;8(8):R157. DOI: 10.1186/gb-2007-8-8-r157

Tewes M, Aktas B, Welt A, Mueller S, Hauch S, Kimmig R, Kasimir-Bauer S. Molecular profiling and predictive value of circulating tumor cells in patients with metastatic breast cancer: an option for monitoring response to breast cancer related therapies. Breast Cancer Res Treat. **2009** Jun;115(3):581-90. DOI: 10.1007/s10549-008-0143-x

Thiery JP, Acloque H, Huang RY, Nieto MA. Epithelial-mesenchymal transitions in development and disease. Cell. **2009** Nov 25;139(5):871-90. DOI: 10.1016/j.cell.2009.11.007

Thim L. A new family of growth factor-like peptides. 'Trefoil' disulphide loop structures as a common feature in breast cancer associated peptide (pS2), pancreatic spasmolytic polypeptide (PSP), and frog skin peptides (spasmolysins). FEBS Lett. **1989** Jun 19;250(1):85-90. PMID: 2737304

Thim L and May FE. Structure of mammalian trefoil factors and functional insights. Cell Mol Life Sci. **2005** Dec;62(24):2956-73. DOI: 10.1007/s00018-005-5484-6

Thompson SL, Bakhoum SF, Compton DA. Mechanisms of chromosomal instability. Curr Biol. **2010** Mar 23;20(6):R285-95. DOI: 10.1016/j.cub.2010.01.034

Tomasetto C, Rio MC, Gautier C, Wolf C, Hareuveni M, Chambon P, Lathe R. hSP, the domain-duplicated homolog of pS2 protein, is co-expressed with pS2 in stomach but not in breast carcinoma. EMBO J. **1990** Feb;9(2):407-14. PMCID: PMC551681

Torres L, Ribeiro FR, Pandis N, Andersen JA, Heim S, Teixeira MR. Intratumor genomic heterogeneity in breast cancer with clonal divergence between primary carcinomas and lymph node metastases. Breast Cancer Res Treat. **2007** Apr;102(2):143-55. DOI: 10.1007/s10549-006-9317-6

Tsai JH, Donaher JL, Murphy DA, Chau S, Yang J. Spatiotemporal regulation of epithelial-mesenchymal transition is essential for squamous cell carcinoma metastasis. Cancer Cell. **2012** Dec 11;22(6):725-36. DOI: 10.1016/j.ccr.2012.09.022

Turner N, Tutt A, Ashworth A. Hallmarks of 'BRCAness' in sporadic cancers. Nat Rev Cancer. **2004** Oct;4(10):814-9. DOI: 10.1038/nrc1457

Turner NC, Ro J, André F, Loi S, Verma S, Iwata H, Harbeck N, Loibl S, Huang Bartlett C, Zhang K, Giorgetti C, Randolph S, Koehler M, Cristofanilli M; PALOMA3 Study Group. Palbociclib in Hormone-Receptor-Positive Advanced Breast Cancer. N Engl J Med. **2015** Jul 16;373(3):209-19. DOI: 10.1056/NEJMoa1505270

Uchino H, Kataoka H, Itoh H, Hamasuna R, Koono M. Overexpression of intestinal trefoil factor in human colon carcinoma cells reduces cellular growth in vitro and in vivo. Gastroenterology. **2000** Jan;118(1):60-9. PMID: 10611154

Ursini-Siegel J, Schade B, Cardiff RD, Muller WJ. Insights from transgenic mouse models of ERBB2-induced breast cancer. Nat Rev Cancer. **2007** May;7(5):389-97. DOI: 10.1038/nrc2127

Valent P, Bonnet D, De Maria R, Lapidot T, Copland M, Melo JV, Chomienne C, Ishikawa F, Schuringa JJ, Stassi G, Huntly B, Herrmann H, Soulier J, Roesch A, Schuurhuis GJ, Wöhrer S, Arock M, Zuber J, Cerny-Reiterer S, Johnsen HE, Andreeff M, Eaves C. Cancer stem cell definitions and terminology: the devil is in the details. Nat Rev Cancer. **2012** Nov;12(11):767-75. DOI: 10.1038/nrc3368

van Amerongen R, Bowman AN, Nusse R. Developmental stage and time dictate the fate of Wnt/β-catenin-responsive stem cells in the mammary gland. Cell Stem Cell. **2012** Sep 7;11(3):387-400. DOI: 10.1016/j.stem.2012.05.023

van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, van der Velde T, Bartelink H, Rodenhuis S, Rutgers ET, Friend SH, Bernards R. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med. **2002** Dec 19;347(25):1999-2009. DOI: 10.1056/NEJMoa021967

Van Keymeulen A, Rocha AS, Ousset M, Beck B, Bouvencourt G, Rock J, Sharma N, Dekoninck S, Blanpain C. Distinct stem cells contribute to mammary gland development and maintenance. Nature. **2011** Oct 9;479(7372):189-93. DOI: 10.1038/nature10573

van Schooneveld E, Wildiers H, Vergote I, Vermeulen PB, Dirix LY, Van Laere SJ. Dysregulation of microRNAs in breast cancer and their potential role as prognostic and predictive biomarkers in patient management. Breast Cancer Res. **2015** Feb 18;17:21. DOI: 10.1186/s13058-015-0526-y

van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH. Gene expression profiling predicts clinical outcome of breast cancer. Nature. **2002** Jan 31;415(6871):530-6. DOI: 10.1038/415530a

Velasco-Velázquez MA, Li Z, Casimiro M, Loro E, Homsi N, Pestell RG. Examining the role of cyclin D1 in breast cancer. Future Oncol. **2011** Jun;7(6):753-65. DOI: 10.2217/fon.11.56

Veronesi U, Banfi A, Saccozzi R, Salvadori B, Zucali R, Uslenghi C, Greco M, Luini A, Rilke F, Sultan L. Conservative treatment of breast cancer. A trial in progress at the Cancer Institute of Milan. Cancer. **1977** Jun;39(6 Suppl):2822-6. PMID: 326380

Veronesi U, Paganelli G, Galimberti V, Viale G, Zurrida S, Bedoni M, Costa A, de Cicco C, Geraghty JG, Luini A, Sacchini V, Veronesi P. Sentinel-node biopsy to avoid axillary dissection in breast cancer with clinically negative lymph-nodes. Lancet. **1997** Jun 28;349(9069):1864-7. DOI: 10.1016/S0140-6736(97)01004-0

Veronesi U, Saccozzi R, Del Vecchio M, Banfi A, Clemente C, De Lena M, Gallus G, Greco M, Luini A, Marubini E, Muscolino G, Rilke F, Salvadori B, Zecchini A, Zucali R. Comparing radical mastectomy with quadrantectomy, axillary dissection, and radiotherapy in patients with small cancers of the breast. N Engl J Med. **1981** Jul 2;305(1):6-11. DOI: 10.1056/NEJM198107023050102

Veronesi U, Cascinelli N, Mariani L, Greco M, Saccozzi R, Luini A, Aguilar M, Marubini E. Twenty-year follow-up of a randomized study comparing breast-conserving surgery with radical mastectomy for early breast cancer. N Engl J Med. **2002** Oct 17;347(16):1227-32. DOI: 10.1056/NEJMoa020989

Viedma-Rodríguez R, Baiza-Gutman L, Salamanca-Gómez F, Diaz-Zaragoza M, Martínez-Hernández G, Ruiz Esparza-Garrido R, Velázquez-Flores MA, Arenas-Aranda D. Mechanisms associated with resistance to tamoxifen in estrogen receptor-positive breast cancer (review). Oncol Rep. **2014** Jul;32(1):3-15. DOI: 10.3892/or.2014.3190

Virchow R. Cellular pathology. As based upon physiological and pathological histology. Lecture XVI--Atheromatous affection of arteries. 1858. Nutr Rev. **1989** Jan;47(1):23-5. PMID: 2649802

Viski C, König C, Kijewska M, Mogler C, Isacke CM, Augustin HG. Endosialin-Expressing Pericytes Promote Metastatic Dissemination. Cancer Res. **2016** Sep 15;76(18):5313-25. DOI: 10.1158/0008-5472.CAN-16-0932

Visvader JE, Lindeman GJ. Cancer stem cells in solid tumours: accumulating evidence and unresolved questions. Nat Rev Cancer. **2008** Oct;8(10):755-68. DOI: 10.1038/nrc2499

Visvader JE and Stingl J. Mammary stem cells and the differentiation hierarchy: current status and perspectives. Genes Dev. **2014** Jun 1;28(11):1143-58. DOI: 10.1101/gad.242511.114

Vogel CL, Cobleigh MA, Tripathy D, Gutheil JC, Harris LN, Fehrenbacher L, Slamon DJ, Murphy M, Novotny WF, Burchmore M, Shak S, Stewart SJ, Press M. Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer. J Clin Oncol. **2002** Feb 1;20(3):719-26. DOI: 10.1200/JCO.2002.20.3.719

Vona G, Sabile A, Louha M, Sitruk V, Romana S, Schütze K, Capron F, Franco D, Pazzagli M, Vekemans M, Lacour B, Bréchot C, Paterlini-Bréchot P. Isolation by size of epithelial tumor cells: a new method for the immunomorphological and molecular characterization of circulating tumor cells. Am J Pathol. **2000** Jan;156(1):57-63. DOI: 10.1016/S0002-9440(10)64706-2

Wan X, Harkavy B, Shen N, Grohar P, Helman LJ. Rapamycin induces feedback activation of Akt signaling through an IGF-1R-dependent mechanism. Oncogene. **2007** Mar 22;26(13):1932-40. DOI: 10.1038/sj.onc.1209990

Wang Y, Waters J, Leung ML, Unruh A, Roh W, Shi X, Chen K, Scheet P, Vattathil S, Liang H, Multani A, Zhang H, Zhao R, Michor F, Meric-Bernstam F, Navin NE. Clonal evolution in breast cancer revealed by single nucleus genome sequencing. Nature. **2014** Aug 14;512(7513):155-60. DOI: 10.1038/nature13600

Wang C, Mu Z, Chervoneva I, Austin L, Ye Z, Rossi G, Palazzo JP, Sun C, Abu-Khalaf M, Myers RE, Zhu Z, Ba Y, Li B, Hou L, Cristofanilli M, Yang H. Longitudinally collected CTCs and CTC-clusters and clinical outcomes of metastatic breast cancer. Breast Cancer Res Treat. **2016** Oct 22. [Epub ahead of print] DOI: 10.1007/s10549-016-4026-2

Weigelt B, Verduijn P, Bosma AJ, Rutgers EJ, Peterse HL, van't Veer LJ. Detection of metastases in sentinel lymph nodes of breast cancer patients by multiple mRNA markers. Br J Cancer. **2004** Apr 19;90(8):1531-7. DOI: 10.1038/sj.bjc.6601659

Weiner DB, Liu J, Cohen JA, Williams WV, Greene MI. A point mutation in the neu oncogene mimics ligand induction of receptor aggregation. Nature. **1989** May 18;339(6221):230-1. DOI:10.1038/339230a0

Welboren WJ, Stunnenberg HG, Sweep FC, Span PN. Identifying estrogen receptor target genes. Mol Oncol. **2007** Sep;1(2):138-43. DOI: 10.1016/j.molonc.2007.04.001

Westley BR, Griffin SM, May FE. Interaction between TFF1, a gastric tumor suppressor trefoil protein, and TFIZ1, a brichos domain-containing protein with homology to SP-C. Biochemistry. **2005** Jun 7;44(22):7967-75. PMID: 15924415

Whittle JR, Lewis MT, Lindeman GJ, Visvader JE. Patient-derived xenograft models of breast cancer and their predictive power. Breast Cancer Res. **2015** Feb 10;17:17. DOI: 10.1186/s13058-015-0523-1

Wicha MS and Hayes DF. Circulating tumor cells: not all detected cells are bad and not all bad cells are detected. J Clin Oncol. **2011** Apr 20;29(12):1508-11. DOI: 10.1200/JCO.2010.34.0026

Wicha MS, Liu S, Dontu G. Cancer stem cells: an old idea--a paradigm shift. Cancer Res. **2006** Feb 15;66(4):1883-90; discussion 1895-6. DOI: 10.1158/0008-5472.CAN-05-3153

Wiesen JF, Young P, Werb Z, Cunha GR Development. Signaling through the stromal epidermal growth factor receptor is necessary for mammary ductal development. **1999** Jan;126(2):335-44. PMID: 9847247

Wilhelm I, Molnár J, Fazakas C, Haskó J, Krizbai IA. Role of the blood-brain barrier in the formation of brain metastases. Int J Mol Sci. **2013** Jan 11;14(1):1383-411. DOI: 10.3390/ijms14011383

Williams ES, Rodriguez-Bravo V, Chippada-Venkata U, De Ia Iglesia-Vicente J, Gong Y, Galsky M, Oh W, Cordon-Cardo C, Domingo-Domenech J. Generation of Prostate Cancer Patient Derived Xenograft Models from Circulating Tumor Cells. J Vis Exp. **2015** Oct 20;(105):53182. DOI: 10.3791/53182

Wilson EB. The Cell in Development and Inheritance. **1896**. The Macmillan Co. New York, New York, USA

Wilson KS, Roberts H, Leek R, Harris AL, Geradts J. Differential gene expression patterns in HER2/neu-positive and -negative breast cancer cell lines and tissues. Am J Pathol. **2002** Oct;161(4):1171-85. DOI: 10.1016/S0002-9440(10)64394-5

Wolff AC, Hammond ME, Hicks DG, Dowsett M, McShane LM, Allison KH, Allred DC, Bartlett JM, Bilous M, Fitzgibbons P, Hanna W, Jenkins RB, Mangu PB, Paik S, Perez EA, Press MF, Spears PA, Vance GH, Viale G, Hayes DF; American Society of Clinical Oncology; College of American Pathologists. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. J Clin Oncol. **2013** Nov 1;31(31):3997-4013. DOI: 10.1200/JCO.2013.50.9984

Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, Collins N, Gregory S, Gumbs C, Micklem G. Identification of the breast cancer susceptibility gene BRCA2. Nature. **1995** Dec 21-28;378(6559):789-92. DOI: 10.1038/378789a0

Workman P, Aboagye EO, Balkwill F, Balmain A, Bruder G, Chaplin DJ, Double JA, Everitt J, Farningham DA, Glennie MJ, Kelland LR, Robinson V, Stratford IJ, Tozer GM, Watson S, Wedge SR, Eccles SA; Committee of the National Cancer Research Institute. Guidelines for the welfare and use of animals in cancer research. Br J Cancer. **2010** May 25;102(11):1555-77. DOI: 10.1038/sj.bjc.6605642

Wright MH, Calcagno AM, Salcido CD, Carlson MD, Ambudkar SV, Varticovski L. Brca1 breast tumors contain distinct CD44+/CD24- and CD133+ cells with cancer stem cell characteristics. Breast Cancer Res. **2008**;10(1):R10. DOI: 10.1186/bcr1855

Wright NA, Hoffmann W, Otto WR, Rio MC, Thim L. Rolling in the clover: trefoil factor family (TFF)-domain peptides, cell migration and cancer. FEBS Lett. **1997** May 19;408(2):121-3. PMID: 9187350

Wyckoff JB, Wang Y, Lin EY, Li JF, Goswami S, Stanley ER, Segall JE, Pollard JW, Condeelis J. Direct visualization of macrophage-assisted tumor cell intravasation in mammary tumors. Cancer Res. **2007** Mar 15;67(6):2649-56. DOI: 10.1158/0008-5472.CAN-06-1823

Xenidis N, Perraki M, Kafousi M, Apostolaki S, Bolonaki I, Stathopoulou A, Kalbakis K, Androulakis N, Kouroussis C, Pallis T, Christophylakis C, Argyraki K, Lianidou ES, Stathopoulos S, Georgoulias V, Mavroudis D. Predictive and prognostic value of peripheral blood cytokeratin-19 mRNA-positive cells detected by real-time polymerase chain reaction in node-negative breast cancer patients. J Clin Oncol. **2006** Aug 10;24(23):3756-62. DOI: 10.1200/JCO.2005.04.5948

Yadav A, Kumar B, Yu JG, Old M, Teknos TN, Kumar P. Tumor-Associated Endothelial Cells Promote Tumor Metastasis by Chaperoning Circulating Tumor Cells and Protecting Them from Anoikis. PLoS One. **2015** Oct 28;10(10):e0141602. DOI: 10.1371/journal.pone.0141602

Yalcin-Ozuysal O, Fiche M, Guitierrez M, Wagner KU, Raffoul W, Brisken C. Antagonistic roles of Notch and p63 in controlling mammary epithelial cell fates. Cell Death Differ. **2010** Oct;17(10):1600-12. DOI: 10.1038/cdd.2010.37

Yang J and Weinberg RA. Epithelial-mesenchymal transition: at the crossroads of development and tumor metastasis. Dev Cell. **2008** Jun;14(6):818-29. DOI: 10.1016/j.devcel.2008.05.009

Yang J, Mani SA, Donaher JL, Ramaswamy S, Itzykson RA, Come C, Savagner P, Gitelman I, Richardson A, Weinberg RA. Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. Cell. **2004** Jun 25;117(7):927-39. DOI: 10.1016/j.cell.2004.06.006

Yin JJ, Selander K, Chirgwin JM, Dallas M, Grubbs BG, Wieser R, Massagué J, Mundy GR, Guise TA. TGF-beta signaling blockade inhibits PTHrP secretion by breast cancer cells and bone metastases development. J Clin Invest. **1999** Jan;103(2):197-206. DOI: 10.1172/JCI3523

Yu M, Stott S, Toner M, Maheswaran S, Haber DA. Circulating tumor cells: approaches to isolation and characterization. J Cell Biol. **2011** Feb 7;192(3):373-82. DOI: 10.1083/jcb.201010021

Yu M, Bardia A, Wittner BS, Stott SL, Smas ME, Ting DT, Isakoff SJ, Ciciliano JC, Wells MN, Shah AM, Concannon KF, Donaldson MC, Sequist LV, Brachtel E, Sgroi D, Baselga J, Ramaswamy S, Toner M, Haber DA, Maheswaran S. Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. Science. **2013** Feb 1;339(6119):580-4. DOI: 10.1126/science.1228522

Zhang H, Cohen AL, Krishnakumar S, Wapnir IL, Veeriah S, Deng G, Coram MA, Piskun CM, Longacre TA, Herrler M, Frimannsson DO, Telli ML, Dirbas FM, Matin AC, Dairkee SH, Larijani B, Glinsky GV, Bild AH, Jeffrey SS. Patient-derived xenografts of triple-negative breast cancer reproduce molecular features of patient tumors and respond to mTOR inhibition. Breast Cancer Res. **2014** Apr 7;16(2):R36. DOI: 10.1186/bcr3640

Zhang J, Qiao X, Shi H, Han X, Liu W, Tian X, Zeng X. Circulating tumor-associated neutrophils (cTAN) contribute to circulating tumor cell survival by suppressing peripheral leukocyte activation. Tumour Biol. **2016** Apr;37(4):5397-404. DOI: 10.1007/s13277-015-4349-3

Zhang QX, Borg A, Wolf DM, Oesterreich S, Fuqua SA. An estrogen receptor mutant with strong hormone-independent activity from a metastatic breast cancer. Cancer Res. **1997** Apr 1;57(7):1244-9. PMID: 9102207

Zhang RD, Fidler IJ, Price JE. Relative malignant potential of human breast carcinoma cell lines established from pleural effusions and a brain metastasis. Invasion Metastasis. **1991**;11(4):204-15. PMID: 1765433

Zhang W, Zhu XD, Sun HC, Xiong YQ, Zhuang PY, Xu HX, Kong LQ, Wang L, Wu WZ, Tang ZY. Depletion of tumor-associated macrophages enhances the effect of sorafenib in metastatic liver cancer models by antimetastatic and antiangiogenic effects. Clin Cancer Res. **2010** Jul 1;16(13):3420-30. DOI: 10.1158/1078-0432.CCR-09-2904

Zhou J, Wulfkuhle J, Zhang H, Gu P, Yang Y, Deng J, Margolick JB, Liotta LA, Petricoin E 3rd, Zhang Y. Activation of the PTEN/mTOR/STAT3 pathway in breast cancer stem-like cells is required for viability and maintenance. Proc Natl Acad Sci U S A. **2007** Oct 9;104(41):16158-63. DOI: 10.1073/pnas.0702596104

Zhou F, Drabsch Y, Dekker TJ, de Vinuesa AG, Li Y, Hawinkels LJ, Sheppard KA, Goumans MJ, Luwor RB, de Vries CJ, Mesker WE, Tollenaar RA, Devilee P, Lu CX, Zhu H, Zhang L, Dijke PT. Nuclear receptor NR4A1 promotes breast cancer invasion and metastasis by activating TGF-β signalling. Nat Commun. **2014** Mar 3;5:3388. DOI: 10.1038/ncomms4388

LIST OF PUBLICATIONS

Peer-reviewed journal articles

Papers related to the PhD project

Did circulating tumor cells tell us all they could? The missed circulating tumor cell message in breast cancer.
Fina E, Reduzzi C, Motta R, Di Cosimo S, Bianchi G, Martinetti A, Wechsler J, Cappelletti V, Daidone MG. Int J Biol Markers 30:e429-33 (2015)

Circulating Biomarkers for Prediction of Treatment Response.
Cappelletti V, Appierto V, Tiberio P, Fina E, Callari M, Daidone MG. J Natl Cancer Inst Monogr. 2015:60-3 (2015)

Gene expression profiling of circulating tumor cells in breast cancer.
Fina E, Callari M, Reduzzi C, D'Aiuto F, Mariani G, Generali D, Pierotti MA, Daidone MG, Cappelletti V. Clin Chem. 61:278-89 (2015)

Other papers

Clinical significance of early changes in circulating tumor cells from patients receiving first-line cisplatin-based chemotherapy for metastatic urothelial carcinoma.
Fina E, Necchi A, Giannatempo P, Colecchia M, Raggi D, Daidone MG, and Cappelletti V. Bladder Cancer 2:395–403 (2016)

In-depth characterization of breast cancer tumor-promoting cell transcriptome by RNA sequencing and microarrays.
Callari M, Guffanti A, Soldà G, Merlino G, Fina E, Brini E, Moles A, Cappelletti V, Daidone MG. Oncotarget 7:976-94 (2016)

Main conference abstracts

E. Fina, L. Cleris, M. Dugo, D. Lecis, G. Morandi, V. Cappelletti, M.G. Daidone. Circulating tumor cells as a source of new metastasis-associated biomarkers in breast cancer. Cell-VIB Symposium: Hallmarks of Cancer, December 11 – 13, 2016, Ghent, Belgium

E. Fina, L. Cleris, C. Reduzzi, G. Morandi, F. D'Aiuto, M. Callari, M.G. Daidone, V. Cappelletti. Molecular characterization and biological role of circulating tumor cells: a window into metastasis biology from breast cancer xenograft models. 15th International Biennial Congress of the Metastasis Research Society, June 28 – July 1, 2014, Heidelberg, Germany

DECLARATION OF AUTHORSHIP

This dissertation is submitted for the degree of Doctor of Philosophy at The Open University (Milton Keynes, UK).

I hereby certify that the thesis I am submitting is entirely my own original work, except for what stated below:

- Studies with animal experimental models were performed in collaboration with L. Cleris from Animal House Facility at INT.
- Microarray experiments, STR profiling, DNA extraction and gel-separation of PCR products for Sanger sequencing were performed in collaboration with L. De Cecco, D. Penso and the technical staff from Functional Genomics Unit at INT.
- Sanger sequencing was performed at Eurofins Genomics (Germany).
- CRISPR/Cas9-mediated knock-out experiments, including interpretation of Sanger sequencing results, and lentiviral-delivered stable knock-down experiments, were performed under the supervision of Dr. D. Lecis from Department of Experimental Oncology and Molecular Medicine at INT.
- Processing of FFPE and OCT-embedded frozen samples, and IHC and histological staining were performed by G. Morandi, from Biomarkers Unit, and by L. Ventura and L. Gioiosa, from IHC Facility at INT.
- Collection and processing by ScreenCell® Cyto kit of blood samples from breast cancer patients were performed by Dr. C. Reduzzi and R. Motta from Biomarkers Unit, in collaboration with Dr. S. Di Cosimo, Dr. G. Galli, A. Martinetti and E. Sottotetti from Medical Oncology Department, at INT.
- Cytopathologic analysis of blood samples from breast cancer patients and mouse xenograft models, processed by ScreenCell® Cyto kit, was performed by Dr. J. Wechsler from ScreenCell (Paris).
- Bioinformatic analyses of gene expression profile data were performed by Dr. M. Dugo from Functional Genomics Unit at INT.

- Collection of clinical data was performed in collaboration with staff from Medical Oncology Department at INT.
- Statistical analyses to correlate circulating tumor cell molecular profile with clinical outcome in breast cancer case series were performed by Dr. M.G. Daidone from Biomarkers Unit at INT.
- Statistical analyses to correlate circulating tumor cell count with clinical data in breast cancer case series were performed by Dr. V. Cappelletti from Biomarkers Unit at INT.

All direct or indirect sources and any use of the works of any other author, in any form, are acknowledged as references.

For the comparison of my work with existing sources I agree that it shall be entered in a database where it shall also remain after examination. Further rights of reproduction and usage, however, are not granted here. The results of my thesis were not previously presented to another examination board, and they have not been published except, where otherwise stated.

ACKNOWLEDGEMENTS

I wish to express my appreciation and gratitude to my mentor, Dr. Maria Grazia Daidone, for giving me the opportunity to conceive and carry out an original research project that matches my academic personality, for her invaluable scientific support during my PhD study and related research, and for her precious advice and guidance during my training in her research group. I am also grateful to my supervisor Dr. Vera Cappelletti, for introducing me to the fascinating field of study on circulating tumor cells, for her precious teaching, and for spending considerable time on clarifying experimental issues, debating the countless scientific aspects related to my research and commenting on the thesis manuscript. I sincerely thank also Dr Robert Clarke, for holding with me insightful scientific discussion, for encouraging me and for providing suggestions to face experimental issues.

I am grateful to all collaborators and the scientific staff, who devoted considerable time to my research with their experience, especially Loredana Cleris and Matteo Dugo, and all of my former and current colleagues at Biomarkers Unit, for providing technical support and for sharing difficult and pleasant moments during these 7 years.

I also wish to thank the former Director of PhD Programmes, Dr. Domenico Delia, and the current Director, Dr. Luca Roz, for their guidance, and all the staff who coordinate the PhD students' activities at our Institute.

Clinical Chemistry 61:1
278–289 (2015)

Gene Expression Profiling of Circulating Tumor Cells in Breast Cancer

Emanuela Fina,[1] Maurizio Callari,[1] Carolina Reduzzi,[1] Francesca D'Aiuto,[1] Gabriella Mariani,[2] Daniele Generali,[4] Marco A. Pierotti,[3] Maria G. Daidone,[1*] and Vera Cappelletti[1]

BACKGROUND: Determining the transcriptional profile of circulating tumor cells (CTCs) may allow the acquisition of clinically relevant information while overcoming tumor heterogeneity-related biases associated with use of tissue samples for biomarker assessment. However, such molecular characterization is challenging because CTCs are rare and outnumbered by blood cells.

METHODS: Here, we describe a technical protocol to measure the expression of >29 000 genes in CTCs captured from whole blood with magnetic beads linked with antibodies against epithelial cell adhesion molecule (EpCAM) and the carcinoma-associated mucin, MUC1, designed to be used for CTC characterization in clinical samples. Low numbers of cells (5–200) from the MCF7 and MDA-MB-468 breast cancer cell lines were spiked in healthy donor blood samples and isolated with the AdnaTest EMT-1/Stem CellSelect kit. Gene expression profiles (GEPs) were obtained with the WG-DASL HT assay and compared with GEPs obtained from RNA isolated from cultured cell lines and unspiked samples.

RESULTS: GEPs from samples containing 25 or more spiked cells correlated ($r = 0.95$) with cognate 100-ng RNA input samples, clustered separately from blood control samples, and allowed MCF7 and MDA-MB-468 cells to be distinguished. GEPs with comparable technical quality were also obtained in a preliminary series of clinical samples.

CONCLUSIONS: Our approach allows technically reliable GEPs to be obtained from isolated CTCs for the acquisition of biologically useful information. It is reproducible and suitable for application in prospective studies to assess the clinical utility of CTC GEPs, provided that >25 CTCs can be isolated.

© 2014 American Association for Clinical Chemistry

Attempting to guide cancer treatment on the basis of the features of the primary tumor (PT)[5] *(1–4)* may not represent an optimal approach owing to frequently reported discordance between the primary tumor and metastatic sites *(5, 6)* and the occurrence of intratumoral clonal heterogeneity *(7, 8)*. Moreover, in clinical practice, metastatic lesions are seldom biopsied owing to their anatomic inaccessibility or the comorbidity associated with the procedure. Blood-based biomarker monitoring represents a new direction in the development of a precision medicine approach tailored to provide information on the specific progression step of the disease.

Circulating tumor cells (CTCs) are the purported intermediates of metastatic dissemination and are likely to contain cellular clones responsible for disease progression; CTCs therefore represent a preferred source for the identification of drug targets. Beyond the clinical validity of CTC counts in both early and metastatic breast cancer patients *(9–11)*, CTCs offer the possibility of obtaining information on the disease in real time without invasive biopsies; a biological characterization of such cells is likely to be more representative of the disease evolution and treatment resistance than the PT *(12–14)*. Unfortunately, molecular characterization of CTCs is seriously hampered by their low numbers and by the contamination of blood samples with leukocytes. Primarily for these 2 reasons, most studies have pursued a candidate gene approach focusing on a limited number of genes known to give therapeutic information *(15–17)*, and only few studies have performed an unbiased characterization of numerous genes by array approaches *(18)*. Because CTCs are likely to represent a very heterogeneous cell population, single-cell profiling has been applied

[1] Biomarkers Unit, Department of Experimental Oncology and Molecular Medicine, [2] Medical Oncology Unit, and [3] Scientific Directorate, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy; [4] U.O. Multidisciplinare di Patologia Mammaria, U.S. Terapia Molecolare e Farmacogenomica, A.O. Istituti Ospitalieri di Cremona, Cremona, Italy.

* Address correspondence to this author at: Department of Experimental Oncology and Molecular Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori, Via G.A. Amadeo 42, 20133, Milan, Italy. Fax +39-02-2390-2674; e-mail mariagrazia.daidone@istitutotumori.mi.it.

Received June 26, 2014; accepted October 31, 2014.
Previously published online at DOI: 10.1373/clinchem.2014.229476
© 2014 American Association for Clinical Chemistry

[5] Nonstandard abbreviations: PT, primary tumor; CTC, circulating tumor cells; EpCAM, epithelial cell adhesion molecule; DASL, cDNA-mediated annealing, selection, extension, and ligation; HDB, healthy donor blood; IRCCS, Istituto di Ricovero e Cura a Carattere Scientifico; GEP, gene expression profile; Ct, cycle threshold.

to investigate their transcriptional heterogeneity *(19–22)* as well as multimarker approaches on the bulk of CTC population *(23–27)*.

With the goal of providing a reliable assay allowing the acquisition of valuable information on CTC features in the clinical setting, we have adapted a commercially available method that captures CTCs by use of beads coated with antibodies against epithelial cell adhesion molecule (EpCAM) and the carcinoma-associated mucin, MUC1. Extensive gene expression profiling is then performed on the captured cells with the cDNA-mediated annealing, selection, extension, and ligation (DASL) platform, which allows measurement of the expression of 29 000 genes in low-quantity RNA samples *(28)*. The method has been tested with cell line cells spiked at different numbers into healthy donor blood (HDB) samples, and has been subsequently validated in metastatic breast cancer patients entering systemic treatment at the Fondazione Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) Istituto Nazionale Tumori of Milan and at the Breast Unit of Cremona Hospital.

Materials and Methods

CELL CULTURE AND CELL SPIKING

Breast cancer cell lines MCF7 and MDA-MB-468 (ATCC) were grown in DMEM/F-12 (Lonza) medium supplemented with 10% fetal bovine serum (Lonza) in humidified 5% CO_2 atmosphere. Cell lines were authenticated by short tandem repeats DNA profiling with the StemElite™ ID System kit (Promega). For spiking experiments, cells were harvested with TrypLE™ Select 1X (Invitrogen, Life Technologies), and single cells from highly diluted cell suspensions were micropipetted under an inverted optical microscope directly into 5 mL of whole HDB in AdnaCollect EDTA collection tubes (AdnaGen). Samples were processed immediately or stored at 4 °C for ≤2 h after spiking.

CLINICAL SAMPLES

Gene expression profile analysis was carried out on CTCs quantified in parallel by different approaches (CellSearch and AdnaTest) from blood of 7 patients with advanced breast cancers entering primary systemic treatment protocols ongoing at the Breast Unit, Istituti Ospitalieri at Cremona (6 cases, CTC counted by CellSearch) and at the Medical Oncology Department at Istituto Nazionale Tumori, Milan (1 case, CTC evaluated by AdnaTest). Blood samples for CTC gene expression (5 mL) were collected before starting a new line of treatment, drawn into BD Vacutainer K_3EDTA tubes, and processed within 1 h after collection. Blood samples (7.5 mL) for CTC counts with the CellSearch® System were collected in parallel into CellSave Preservative sample tubes (Janssen Diagnostic), whereas an additional 5 mL were collected in parallel for CTC determination by AdnaTest. Cremona and Milan institutional review board and ethics committees approved the study protocol, and all patients provided written informed consent.

The patients' clinicopathologic features and CTC status/counts are summarized in Supplemental Table 1, which accompanies the online version of this article at http://www.clinchem.org/content/vol61/issue1.

TUMOR CELL ENRICHMENT, RNA EXTRACTION, AND SINGLE-GENE EXPRESSION EVALUATION

We isolated tumor cells spiked into blood by the AdnaTest EMT-1/StemCell Select kit (AdnaGen AG). Captured cells were incubated with lysis buffer from the Agencourt RNAdvance Cell v2 kit (Beckman Coulter) for 30 min at room temperature, and bead-free lysates were stored at −80 °C. We performed total RNA extraction according to the manufacturer's instructions; RNA was eluted in 5 or 10 μL nuclease-free water (Ambion), depending on the experimental setting.

For single-gene expression studies, enriched tumor cells were incubated with the AdnaTest lysis buffer, and mRNA was isolated with the Dynabeads® mRNA Direct™ Micro Kit included in the AdnaTest and reverse-transcribed. We used the cDNA obtained as template for evaluation of *PTPRC* (protein tyrosine phosphatase, receptor type, C)[6] (the gene coding for CD45) with the TaqMan assay Hs 00365634_g1 (Applied Biosystems).

WHOLE GENOME EXPRESSION PROFILING ASSAY

We performed the Illumina Human Whole-Genome DASL HT Assay according to manufacturer's instructions. The BeadChips were imaged on the BeadArray Reader.

A flow chart summarizing the experimental design is reported in Fig. 1.

DATA PROCESSING AND STATISTICS

Microarray raw data were generated with Illumina BeadStudio 3.8 software and processed with the Lumi package *(29)* of Bioconductor. For each gene, we selected the probe with the highest interquartile range. Array data were deposited at Gene Expression Omnibus data repository (GSE55470).

[6] Human genes: *PTPRC,* protein tyrosine phosphatase, receptor type, C; *EPCAM,* epithelial cell adhesion molecule; *MUC1,* mucin 1, cell surface associated; *ERBB2,* v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 2, (HER2); *AKT2,* v-akt murine thymoma viral oncogene homolog 2; *PIK3CA,* phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha; *TWIST1,* twist family bHLH transcription factor 1; *ALHD1A1,* aldehyde dehydrogenase 1 family, member A1. See Fig. 6 legend for gene symbols and names mentioned only in that figure.

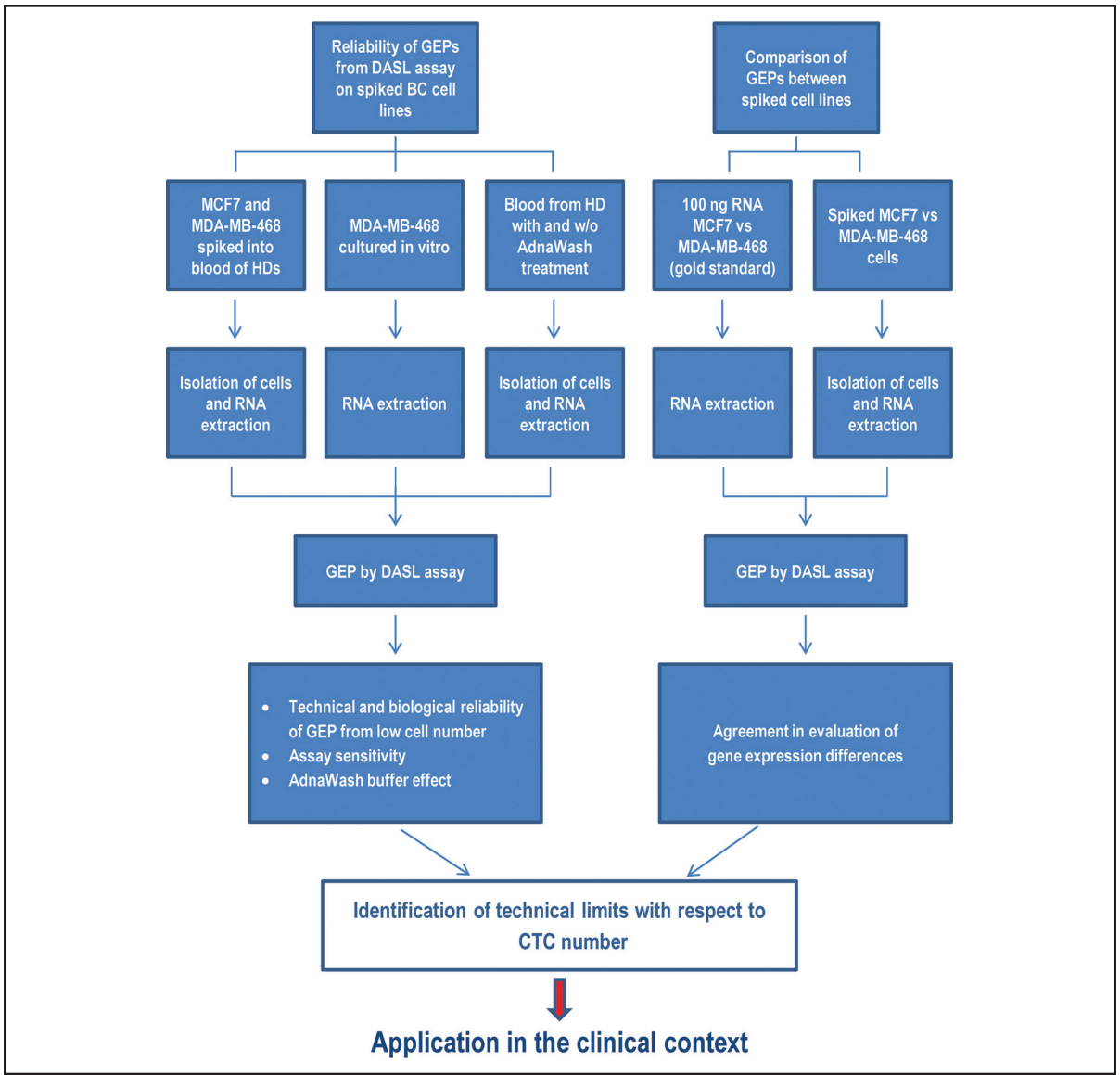


Fig. 1. Flowchart summarizing the experimental design.

The experimental steps carried out on isolated RNA extracted from cell lines and from healthy donor blood samples spiked with different amounts of cells are reported highlighting for each step the technical and biological issues addressed by the experiment. BC, breast cancer; HD, healthy donor.

Unsupervised hierarchical clustering was performed with the Geneplotter package with Euclidean distance and average linkage.

Results and Discussion

SUITABILITY OF THE DASL ASSAY TO DETERMINE GENE EXPRESSION PROFILES OF EPITHELIAL CELLS SPIKED AT LOW FREQUENCIES IN HDB AND EFFICACY OF AdnaWash BUFFER

To evaluate the reliability of the DASL assay for obtaining gene expression profiles (GEPs) from both low RNA-input samples and CTCs, experiments were run with low predefined numbers of breast cancer cells (MCF7 and MDA-MB-468), spiked into a HDB at final concentrations of 400, 200, and 100 cells/5 mL. For this step, a single healthy donor was chosen to ensure comparable samples when investigating the efficiency of leukocyte removal and the interference by leukocyte-specific genes in GEPs from spiked cells.

Total RNA from captured cells was split in 2 equal aliquots and processed in duplicate on 2 distinct chips of the DASL platform together with total RNA (0.5–

100 ng) isolated from the same in vitro cultured cells, in the same amplification/hybridization session. This experiment was designed to (*a*) verify the sensitivity of the assay; (*b*) test the reliability of GEP obtained with 200, 100, and 50 cells; and (*c*) assess the interassay variability (under conditions in which no interference by capture-associated variability was present). In addition, blood samples from the same healthy donor, not containing any spiked cells, were processed similarly with and without use of the AdnaWash buffer (designed to limit leukocyte contamination in the captured samples) to test its ability to prevent contamination by leukocyte-expressed genes.

We generated diagnostic plots (Fig. 2) to judge the technical reliability of the data computing mean signal intensities and probe detection rates (Fig. 2A). As expected, the lowest mean log2 signal intensities (around 5) and probe detection rates (30%) were observed in healthy controls when captured cells were treated with AdnaWash buffer before lysate preparation. These results suggest that after treatment with the washing buffer, few cells remained attached to the immunomagnetic beads. In fact, both the mean signal values and detection rates increased, respectively, to 6% and 45% compared with the lower values obtained in untreated control samples, demonstrating that AdnaWash buffer effectively reduced leukocyte contamination.

The plots in Fig. 2A indicate that both the spiked samples and the samples with higher RNA input were characterized by mean log2 intensities between 7 and 7.7 and probe detection rates >60%. In the case of isolated RNA samples, only those containing 1 or 0.5 ng of RNA showed lower detection rates and intensities.

Fig. 2B summarizes the correlations between samples. Samples from HDB, either untreated or treated with AdnaWash, were weakly correlated with the remaining samples, with a poor correlation between technical duplicates. A separation on the basis of GEP was obtained between the MCF7 and MDA-MB-468 blood-spiked cells, supporting the biological reliability of the GEPs obtained from cells spiked into HDB and captured with immunomagnetic beads. Of note, MDA-MB-468 cells spiked into blood and samples containing their isolated RNA (in the concentration range 10–100 ng) clustered together. However, low-input RNA samples (1 or 0.5 ng) deriving from MDA-MB-468 cells clustered separately. In fact, when the RNA input was decreased from 10 to 1 or 0.5 ng, the correlation coefficients with profiles obtained from a 100-ng RNA sample (an RNA input within the range suggested by the chip manufacturer) dropped from $r = 0.99$ to $r = 0.89$ and $r = 0.83$, respectively. As reported in online Supplemental Fig. 1A, the expression levels of a subset of genes spanning low to intermediate intensities in the high-RNA-input sample were found to be at background levels in samples with lower RNA input. About one-third of the genes were present only in samples with high RNA content, whereas the remainder were found also in low-RNA-input samples. Importantly, the 6578 probes exclusively detected in high-RNA-input samples represent a group of low-expression genes that were not significantly enriched in any specific biological process (see online Supplemental Fig. 1B), suggesting that a lower RNA input affects low-expression genes belonging to any cellular process, without causing any modification in the biological interpretation of data. Such a result, therefore, allows confidence in the GEPs obtained from samples with lower RNA input, as also suggested by April et al. *(28)*.

The efficacy of AdnaWash in removing the contribution of leukocyte-derived genes is further shown in Fig. 2C. Removal of leukocytes is crucial for wide gene expression analysis, since many genes are expressed by both epithelial and mesenchymal cells, and consequently leukocyte contamination can bias GEPs of putative CTCs.

The top scatterplot in Fig. 2C correlates GEPs obtained from the same HDB with or without washing steps with AdnaWash buffer. A substantial number of genes with a wide expression range in samples not treated with AdnaWash appeared to be expressed at low levels after leukocyte removal and, overall, samples not treated with AdnaWash were characterized by higher expression levels for many genes. These data suggest high leukocyte contamination in immunomagnetic bead–captured samples not submitted to AdnaWash treatment.

The next 2 scatterplots in Fig. 2C report correlations between GEPs obtained from a healthy donor sample after a washing step with AdnaWash without any spiked cells and samples from the same donor that were spiked with 50 MCF7 or MDA-MB-468 cells. It appears that for the majority of genes, cell-spiked HDB samples showed higher gene expression levels, suggesting an efficient removal of leukocytes without affecting epithelial cells. Furthermore, a distinct set of genes characterized by a wide range of expression in samples spiked with cells showed only a low expression in unspiked HDB samples, suggesting an enrichment in epithelial-specific genes. The specificity of AdnaWash buffer in the removal of leukocytes is finally supported by the last scatterplot in Fig. 2C, which shows that distinguishing cell-type specific genes from genes equally expressed by MCF7 and MDA-MB-468 cells is possible. On the basis of all the reported observations, we concluded that the treatment with AdnaWash buffer represents a useful step to limit leukocyte contamination that does not affect detection of epithelial-specific genes.

Finally, reproducibility of the AdnaWash procedure for the removal of leukocytes was tested by evaluating the levels of *PTPRC* gene expression with a quantitative PCR approach in 7 samples from the same healthy donor spiked with breast cancer cells. Different numbers of breast cancer cells (0–50) were spiked into 7 independent

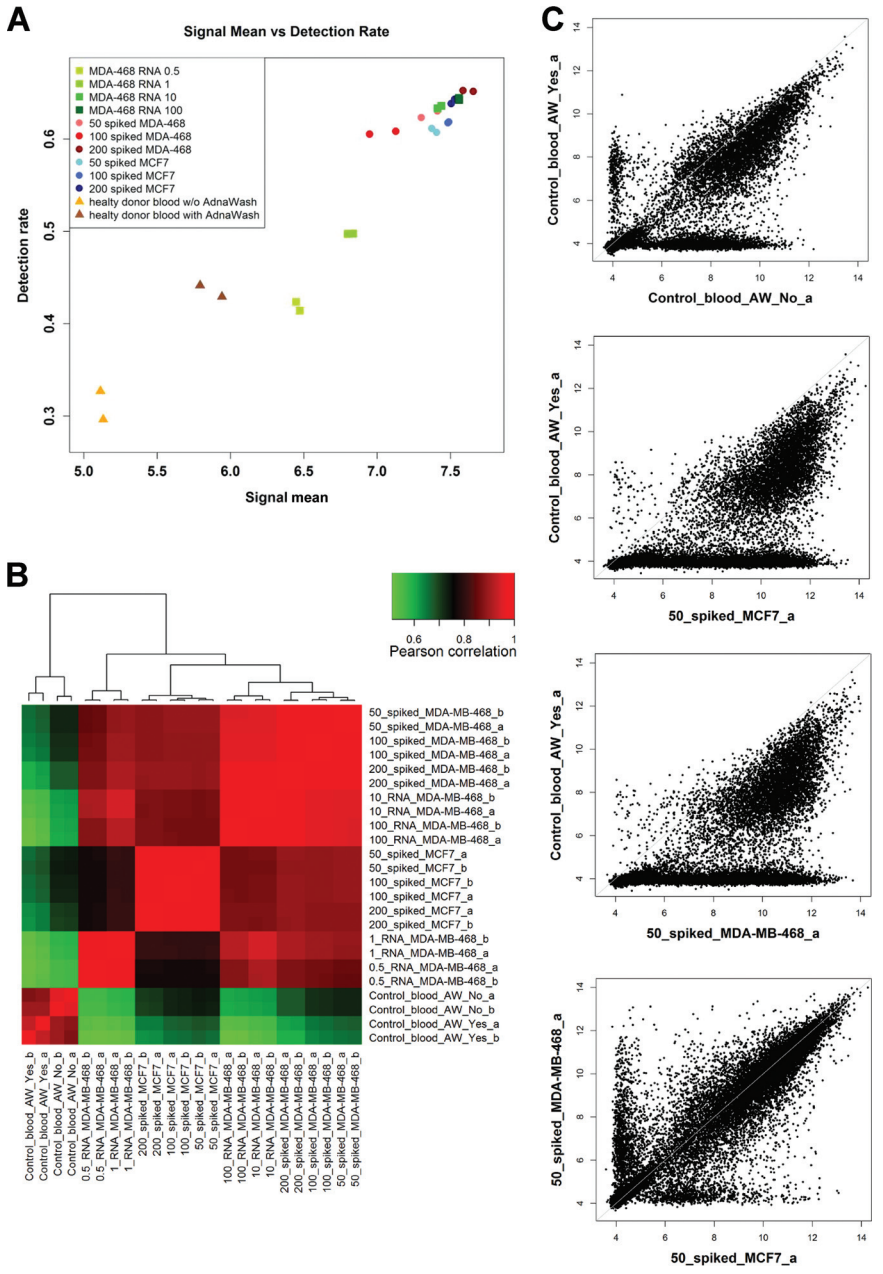


Fig. 2. Intensities and correlations.

(A), Distribution plot of signal mean intensities in gene expression data derived from isolated MDA-MB-468 RNAs (green squares), from 50–100–200 MDA-MB-468 (red dots) or MCF7 (blue dots) cells spiked into healthy donor blood (HDB) and from cell-free HDB (yellow triangles), profiled with the Illumina Human Whole-Genome DASL HT. (B), Heat map reporting the reciprocal correlations (Pearson correlation coefficient) among GEPs obtained from RNA directly isolated from MDA-MB-468 cells or from captured cell RNA corresponding to 200, 100, and 50 MDA-MB-468 or MCF7 cells spiked into HDB. All GEPs were obtained by Illumina Human Whole-Genome DASL HT. Before computation of the correlations, data were separately normalized with the robust spline normalization method. (C), Correlations between samples: AdnaWash buffer–treated vs –untreated HDB samples; AdnaWash buffer–treated HDB containing 50 MCF7-spiked cells vs AdnaWash buffer–treated HDB; AdnaWash buffer–treated HDB containing 50 MDA-MB-468 –spiked cells vs AdnaWash buffer–treated HDB; AdnaWash buffer–treated HDB containing 50 MDA-MB-468 –spiked cells vs AdnaWash buffer–treated HDB containing 50 MCF7-spiked cells. AW, AdnaWash.

5-mL blood samples obtained from the same healthy donor at a single time point to minimize leukocyte variation. After the standard washing procedure with the AdnaWash buffer, captured cells were lysed, and *PTPRC* gene expression was assayed by quantitative PCR. The mean cycle threshold (Ct) value was 36.53 (0.46) with a CV of 1.26%. The CV value obtained supports the reproducibility of the leukocyte removal.

BIOLOGICAL RELIABILITY OF GEPs OBTAINED FROM CELLS SPIKED INTO BLOOD

We assessed the biological reliability of GEPs by comparing genes shared by captured cells and their corresponding isolated RNA and genes selectively expressed by samples derived from isolated RNA (100 ng) or from captured cells (200 cells). Results for MDA-MB-468 cells are shown in Fig. 3A. Reliability of profiles was supported by the high number of shared genes (90.4%) and by the fact that genes exclusively expressed by captured cells (5.5%) were enriched in gene ontology terms referring to leukocytes (e.g., immune response, immune system process, lymphocyte activation), whereas genes exclusively expressed in isolated RNA samples did not show a specific type of gene ontology enrichment. A list of gene ontology terms exclusively enriched in captured cells and in the corresponding isolated RNA is reported in online Supplemental File 1.

The actual enrichment in specific gene categories in cells captured from spiked blood samples compared with control HDB was evaluated considering both types of controls, either washed or not with AdnaWash buffer (Fig. 3B). A proportion of genes (41.5%) expressed by all 3 samples was enriched in terms referring to biological functions common to all types of cells (e.g., RNA processing, cellular macromolecule catabolic process, RNA binding), whereas a smaller set of genes (21.9%) was shared between samples spiked with cells and the control samples that were not treated to remove leukocytes. Such genes that are enriched in cellular biological functions common to all types of cells likely represent common genes with lower expression levels, and this could explain their absence in the AdnaWash-treated samples. The efficacy of leukocyte removal was supported by the Gene Ontology of the 430 genes exclusively expressed in controls untreated with AdnaWash that were enriched in 151 gene ontology terms that all referred to leukocytes and included immune response, inflammation, leukocyte activation, macrophage activation, and lymphocyte differentiation (see online Supplemental File 2). The same analysis was repeated for MCF7-spiked cells, and a similar interpretation could be made (see online Supplemental Fig. 2 and Supplemental File 3). Similar results were also obtained by repeating the analysis with the replicated samples (data not shown).

SENSITIVITY AND EXPERIMENTAL VARIABILITY OF THE DEVELOPED CTC GENE PROFILING METHOD

The sensitivity and experimental variability of the entire CTC capture and profiling assay was tested by spiking independent triplicates of 50, 25, 10, and 5 MCF7 and MDA-MB-468 cells into HDB. RNA (100 ng) extracted from the 2 cell lines was also hybridized in triplicate as a gold-standard sample, and in the case of the MCF7 cell line, triplicate spikes with 50 cells were also performed in parallel in a different donor.

To understand the quality of the data, we carried out Pearson correlation analysis on all samples. The resulting correlation plot is shown in Fig. 4. One sample (10 spiked MDA-MB-468 cells) found to be completely unrelated to the other samples was rated as a technical failure and was not taken into consideration in the following analyses (see also online Supplemental Fig. 3).

The remaining samples segregated into 2 clusters. A larger cluster (cluster A in Fig. 4) contained all samples derived from isolated RNA and 69% (18/26) of spiked cell samples. In both the RNA samples cluster and the spiked cells cluster, a clear separation between MCF7 and MDA-MB-468 cells was observed. For both MCF7 and MDA-MB-468 cell lines, 73% and 63.6% of spiked samples, respectively, correlated as expected.

The specificity of capture of CTCs and the efficacy of leukocyte removal is also supported by the strong correlation between GEPs derived from 50 MCF7 cells spiked into different HDB.

A second cluster of less well-correlated samples (cluster B) contained the 3 negative samples (HDB with no spiked cells) and 8 spiked samples (both with MFC7 and with MDA-MB-468 cells) with low cell numbers (4 samples spiked with 5 cells, 3 with 10 cells, and 1 with 25 cells). Although the overall number of spiked cell samples processed was not high, the data suggest that in samples containing 50 CTCs reliable GEPs can be obtained, whereas for samples with 10 or fewer CTCs, the failure rate exceeds the success rate. In fact, the majority of low-input samples were more strongly correlated with donors rather than with other samples containing the same cell type. Finally, profiling of 25 captured cells appears to be feasible, although with a success rate of around 80% (5/6). Considering the low numbers of CTCs usually recovered in clinical samples, we therefore suggest that to ensure the successful profiling of CTCs in clinical samples, a larger volume of blood (10–15 mL) should be used.

In GEPs of CTCs from patients, the main aim is to detect biologically significant differences among individuals. To understand whether the developed method was suitable for this purpose, fold changes in gene expression between the luminal MCF7 cell line and the triple-negative MDA-MB-468 cell line were computed. The quality of the data obtained from spiked cells was assessed by correlating fold changes obtained from 50, 25, and 10 spiked cells with

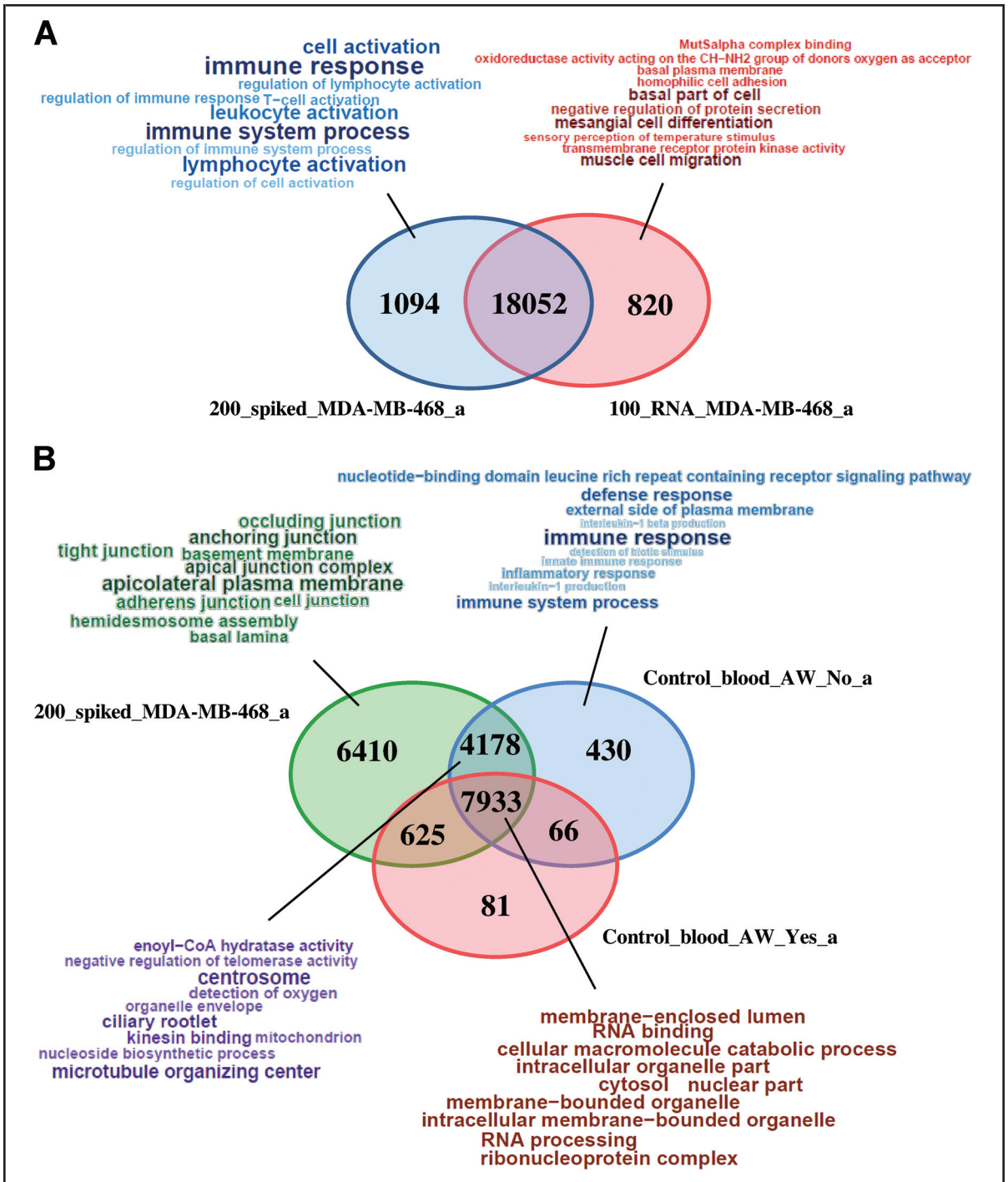


Fig. 3. Gene Ontology.

(A), Eulero-Venn diagrams highlighting the numbers of common and exclusive genes detected in samples derived from 200 spiked MDA-MB-468 cells and 100 ng isolated RNA derived from MDA-MB-468 cells. Gene Ontology analysis results for exclusive genes in each set are reported with character sizes and color intensities directly proportional to the *P* values for the gene set enrichment. (B), Eulero-Venn diagrams highlighting the numbers of common and exclusive genes detected in samples derived from 200 MDA-MB-468 cells spiked into HDB, HDB not treated with AdnaWash buffer, and HDB treated with AdnaWash buffer. Gene ontologies analysis results for exclusive genes in each set are reported with character sizes directly proportional to the gene set enrichment.

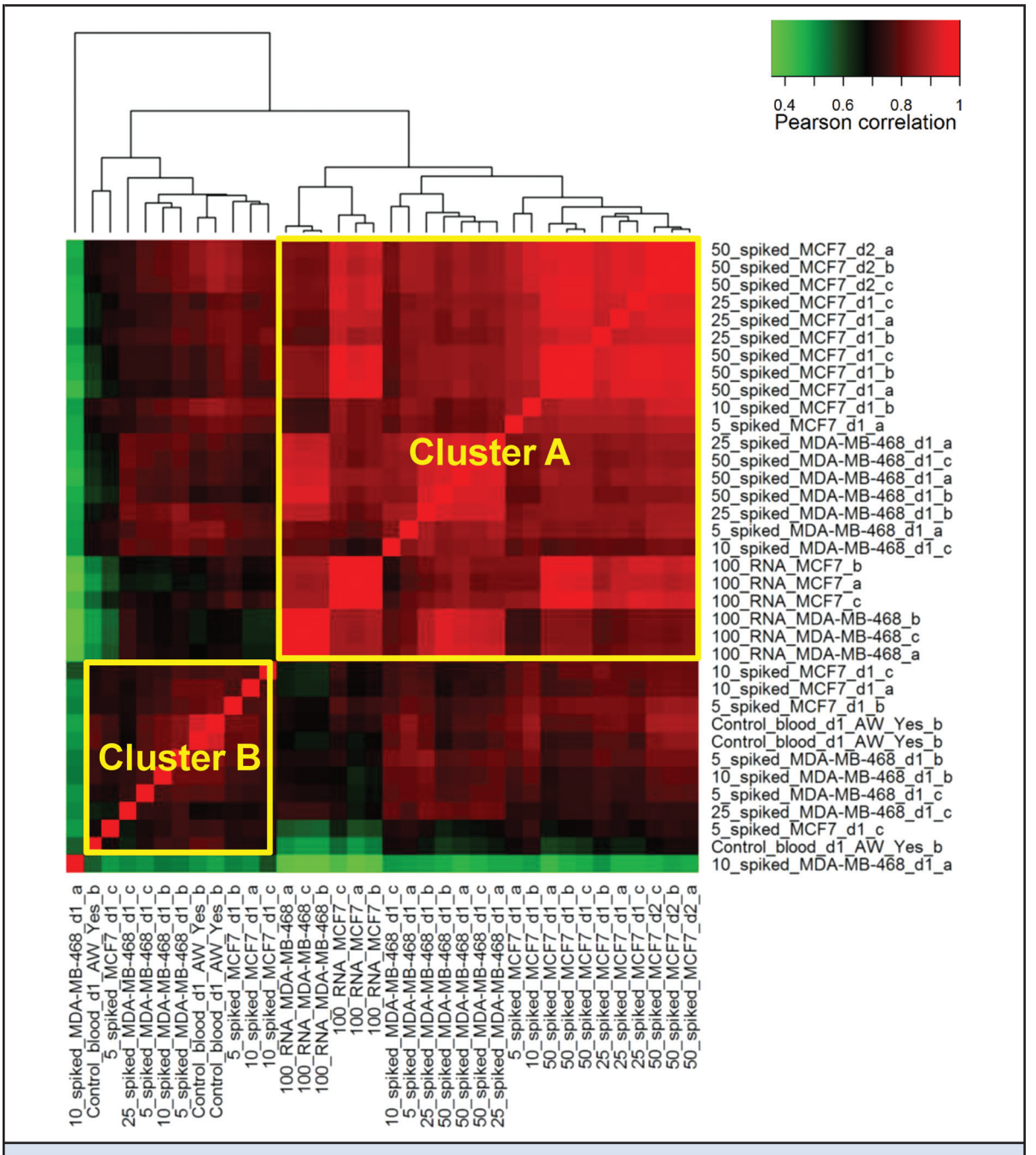


Fig. 4. Heat map reporting the reciprocal correlations (Pearson correlation coefficient) among GEPs obtained from 100 ng RNA directly isolated from MCF7 and MDA-MB-468 cells, or from RNA extracted from 50, 25, 10, and 5 MCF7 or MDA-MB-468 cells spiked in independent triplicates into HDB.

All GEPs were obtained by Illumina Human Whole-Genome DASL HT. Before computation of the correlation values, data were separately normalized with the robust spline normalization method. On the basis of correlations, samples separated in 2 clusters: cluster A containing samples derived from isolated RNA and samples with higher number of spiked cells, and cluster B containing controls and samples spiked with low cell numbers. AW, AdnaWash.

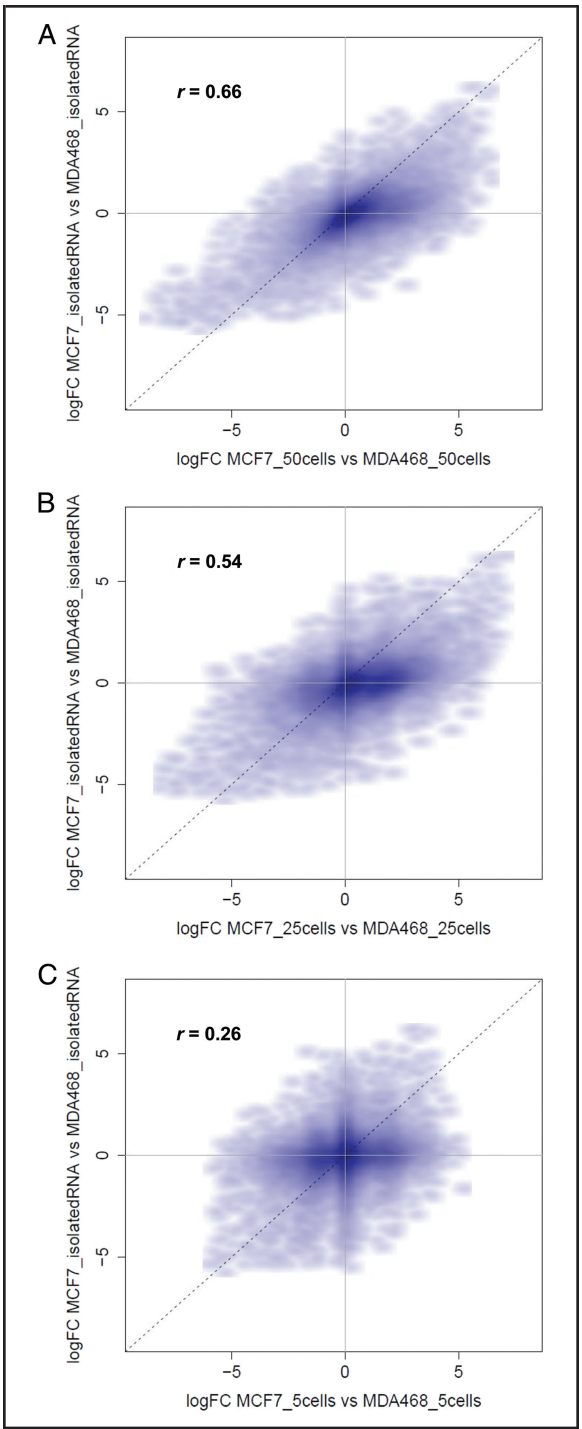


Fig. 5. Paired plot showing correlations between fold changes (FC) for genes found to be differentially expressed between MCF7 cells and MDA-MB-468 cells when 50 (A), 25 (B), and 5 (C) cells were captured from HDB versus FC obtained for differentially expressed genes between isolated RNA samples from MCF7 and MDA-MB-468 cells representing the benchmark.

those obtained with gene expression data derived from RNA (100 ng) directly isolated from cultured cells. Data are reported in Fig. 5. Correlations among log2 fold changes dropped significantly when the number of spiked cells was reduced (fold-change correlations by Pearson coefficient: $r = 0.66$ for 50 cells, $r = 0.54$ for 25 cells, and $r = 0.26$ for 5 cells).

APPLICATION TO CLINICAL SAMPLES

The approach developed for spiked samples was applied to 7 clinical samples. A CellSearch-based CTC number estimation was available for 6 samples, whereas in 1 sample CTC status was evaluated by AdnaTest EMT-1/Stem CellDetect kit without any enumeration. CTC numbers ranged from 0 to 200 cells (in 5 mL whole blood). It is important to underline that this is only an estimation of the true number of CTCs profiled for gene expression, as the profiled CTCs were captured with a different kit. Although the CellSearch approach and the AdnaTest are both based on EpCAM expression, the AdnaTest also contains antibodies directed against cell-surface MUC1 antigen. In the literature, an agreement of 69% and 73% is reported between the 2 methods when using the 5 or 2 CTC/7.5 mL cutoffs *(30)*, respectively.

A heat map for samples clustered on the basis of the expression of the PAM50 gene panel *(31)* is reported in Fig. 6. Except for a sample with a very high number of CTCs (sample CTC001; 200 CTCs in 5 mL blood), no correlation was observed between number of CTCs estimated by CellSearch and the expression levels of PAM50 genes. However, it should be noted that CTC counts were quite low (median CTC count 13). Moreover, the CTCs captured and profiled in this still-preliminary series of clinical samples were obtained from patients characterized by primary tumors of similar subtypes, which does not allow us to observe molecular subtype–related differences. An interesting observation was noted for sample CTC003, in which CTC status was defined as negative by AdnaGen assay on the basis of the low expression of 7 selected genes [*EPCAM* (epithelial cell adhesion molecule), *MUC1* (mucin 1, cell surface associated), *ERBB2* (v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 2, [HER2]), *AKT2* (v-akt murine thymoma viral oncogene homolog 2), *PIK3CA* (phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha), *TWIST1* (twist family bHLH transcription factor 1), and *ALDH1A1* (aldehyde dehydrogenase 1 family, member A1)], but still showed an expression for PAM50 panel genes not dissimilar from that of true CTC-positive samples.

Conclusion

CTCs are rapidly moving from a simple tumor burden marker to a potential biomarker providing information on tumor heterogeneity and tumor biology *(32)*. A recent pooled analysis of patient data has confirmed that CTC

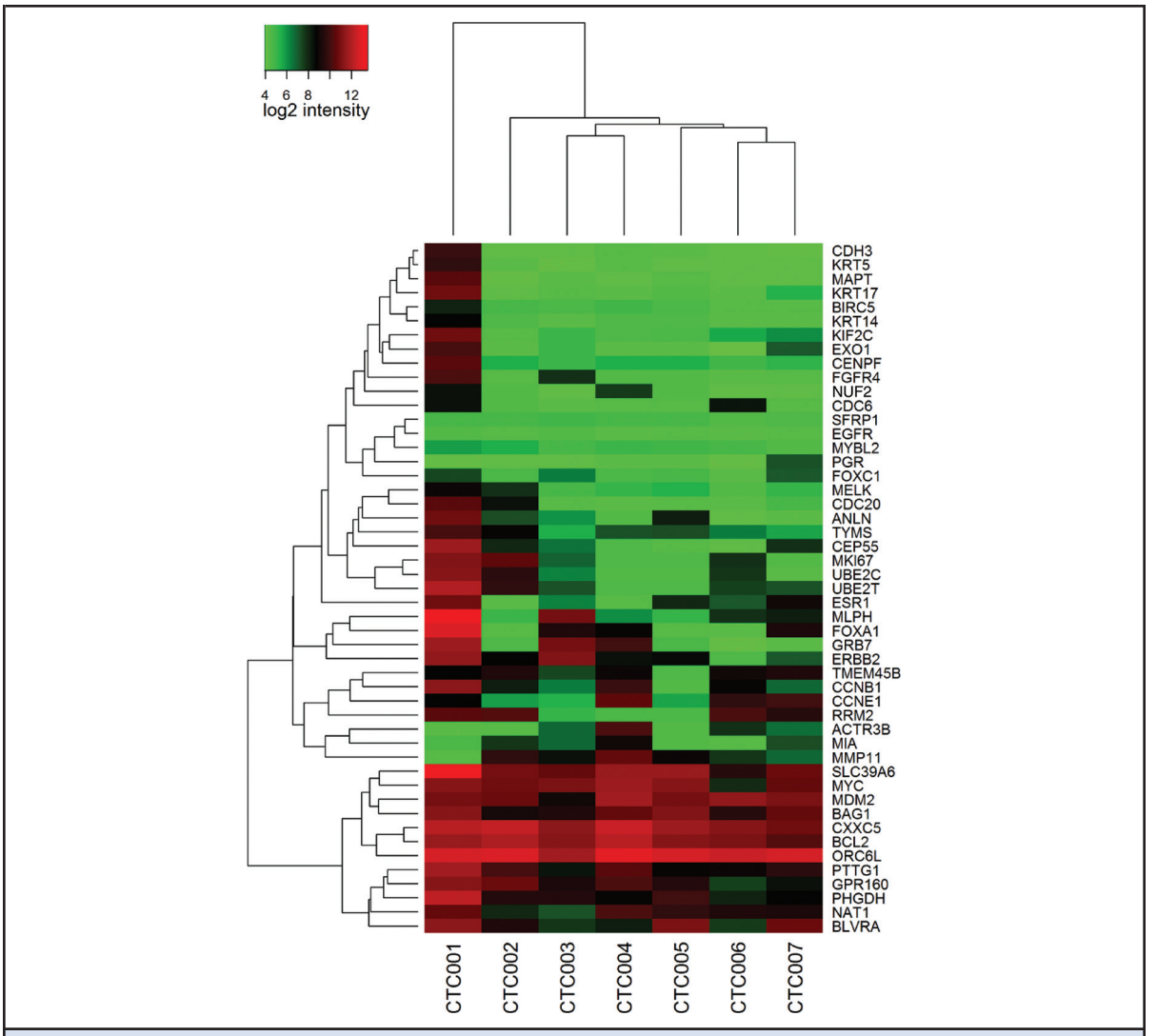


Fig. 6. Unsupervised hierarchical cluster analysis of gene expression profiles from 7 metastatic patient blood samples enriched for CTCs with the AdnaTest EMT-1/Stem CellSelect kit.

Robust spline normalization was applied, and probes with a detection $P < 0.01$ in at least 1 sample were selected. Horizontal rows represent genes from the PAM50 panel, and vertical columns correspond to clinical samples. *CDH3,* cadherin 3, type 1, P-cadherin (placental); *KRT5/17//14,* keratin 1/17/14; *MAPT,* microtubule-associated protein tau; *BIRC5,* baculoviral IAP repeat containing 5; *KIF2C,* kinesin family member 2C; *EXO1,* exonuclease 1; *CENPF,* centromere protein F, 350/400kDa; *FGFR4,* fibroblast growth factor receptor 4; *NUF2,* NDC80 kinetochore complex component (formerly *CDCA1*); *CDC6/20,* cell division cycle 6/20; *SFRP1,* secreted frizzled-related protein 1; *EGFR,* epidermal growth factor receptor; *MYBL2,* v-myb avian myeloblastosis viral oncogene homolog-like 2; *PGR,* progesterone receptor; *FOXC1/A1,* forkhead box C1/A1; *MELK,* maternal embryonic leucine zipper kinase; *ANLN,* anillin, actin binding protein; *TYMS,* thymidylate synthetase; *CEP55,* centrosomal protein 55kDa; *MKI67,* marker of proliferation Ki-67; *UBE2C,* ubiquitin-conjugating enzyme E2C; *UBE2T,* ubiquitin-conjugating enzyme E2T (putative); *ESR1,* estrogen receptor 1; *MLPH,* melanophilin; *GRB7,* growth factor receptor-bound protein 7; *TMEM45B,* transmembrane protein 45B; *CCNB1/E1,* cyclin B1/E1; *RRM2,* ribonucleotide reductase M2; *ACTR3B,* ARP3 actin-related protein 3 homolog B (yeast); *MIA,* melanoma inhibitory activity; *MMP11,* matrix metallopeptidase 11 (stromelysin 3); *SLC39A6,* solute carrier family 39 (zinc transporter), member 6; *MYC,* v-myc avian myelocytomatosis viral oncogene homolog; *MDM2,* MDM2 proto-oncogene, E3 ubiquitin protein ligase; *BAG1,* BCL2-associated athanogene; *CXXC5,* CXXC finger protein 5; *BCL2,* B-cell CLL/lymphoma 2; *ORC6L,* origin recognition complex, subunit 6; *PTTG1,* pituitary tumor-transforming 1; *GPR160,* G protein-coupled receptor 160; *PHGDH,* phosphoglycerate dehydrogenase; *NAT1,* N-acetyltransferase 1 (arylamine N-acetyltransferase); *BLVRA,* biliverdin reductase A.

enumeration is an independent prognostic marker of overall survival and progression-free survival that provides an early assessment of treatment response in breast cancer *(9, 33)*. It is, however, worth reflecting on the advantages offered by CTC molecular characterization in the clinical setting where, so far, the most common approach has been dealing with determination of HER2 status on CTCs.

A direct involvement of CTCs in the metastatic cascade is unquestionable, but probably not all CTC subpopulations have the same metastatic potential. Although not accepted by all *(34)*, epithelial-mesenchymal transition has been suggested to be crucial for dissemination of tumor cells to distant organs *(35)*. Yu et al. *(36)* demonstrated the occurrence of dynamic changes in epithelial and mesenchymal composition of CTCs in patients with metastatic breast cancer and found an association between treatment resistance and presence of CTCs with mesenchymal features in patients who were serially monitored through disease progression. Such a study provides proof of concept that variation of specific CTC subpopulations can serve as an indicator of the rate of cancer adaptation during treatment and represents, therefore, an opportunity to identify new therapeutic targets and resistance markers possibly anticipating clinical progression.

This concept is further reinforced in the light of the well-known intratumor spatial/temporal heterogeneity of both primaries and metastases, which can be overcome by exploiting liquid biopsies *(37)* containing materials (e.g., circulating tumor DNA, CTCs, circulating microRNAs) shed in the bloodstream by primary tumors, but also by the complete set of metastases.

To fully exploit the opportunity offered by CTC molecular characterization in the clinical setting, it therefore is mandatory to have an approach allowing an extensive molecular characterization of CTCs, enabling collection of data that offer timely information on tumor progression and/or development of treatment resistance and are suitable for clinical studies in a multicenter setting.

Despite the many challenges, primarily having to do with low sample input and leukocyte contamination, the experiments run on HDB definitely support the possibility of obtaining technically reliable gene expression profiles from as few as 25 cells in 5 mL blood. Our data also show that the obtained profiles convey biologically useful information that may allow differences among samples to be distinguished.

Several published studies have reported gene expression from CTCs, but these have been limited to tens or hundreds of genes primarily detected by PCR. Whereas some of the studies *(23)* used an EpCAM-based enrichment approach, other studies, mostly dealing with single cells in prostate cancer, relied on complex techniques that are not suitable for translation to the clinic *(38)*.

Our study has the advantage of providing a pipeline for obtaining biologically reliable profiles in a way that is easily applicable to the clinical context; this contrasts with single-cell profiling methods that are not yet ready to be transferred into the daily routine. A limitation in the developed method is represented by the need of somewhat high numbers of cells compared with the median values of CTCs identified in peripheral blood from breast cancer patients. However, such a problem can be solved by drawing higher volumes of blood and improving the CTC-enrichment methods to avoid losses of not fully epithelial CTCs. In fact, on the basis of preliminary data obtained in our laboratory on a limited set of breast cancer patients (21 with early and 9 with metastatic disease), the use of improved antibody cocktails (AdnaTest EMT-1/Stem vs Adna Test EMT-2/Stem CellSelect kits) for CTC enrichment raised the positivity percentage 4- and 3-fold in early and metastatic breast cancer, respectively. Although this does not represent a direct cell count like those obtainable with the CellSearch approach, the increased positivity percentage is likely to be associated with an increase in the absolute number of CTCs. So far, 5 CTCs/7.5 mL is considered the clinically significant CTC threshold for breast cancer patients; even at the lowest end of the CTC number distribution in breast cancer patients, a blood draw of 38 mL would be enough to guarantee the critical number of 25 CTCs necessary for our gene expression approach. Therefore, when using the conventional CTC-enrichment methods, our GEP protocol appears to be more suitable for application in the metastatic context where the number of CTCs is higher than it is in early breast cancer. The AdnaTest-based CTC-capture approach, with its improved antibody cocktail, offers an effective means to increase the number of detectable CTCs.

Author Contributions: *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.*

Authors' Disclosures or Potential Conflicts of Interest: *Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:*

Employment or Leadership: None declared.
Consultant or Advisory Role: None declared.
Stock Ownership: None declared.
Honoraria: None declared.
Research Funding: This work was supported by the European Commission under the 7th Framework Programme under grant agreement no. 260791, EurocanPlatform, and Italian Health Ministry. M.A. Pierotti, AIRC grant 5/1000 #12162; M.G. Daidone, AIRC grant IG #10611.
Expert Testimony: None declared.
Patents: None declared.

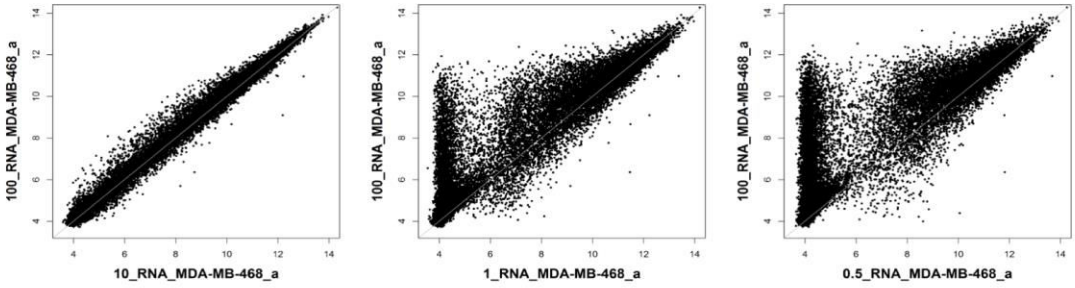
Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, or preparation or approval of manuscript.

References

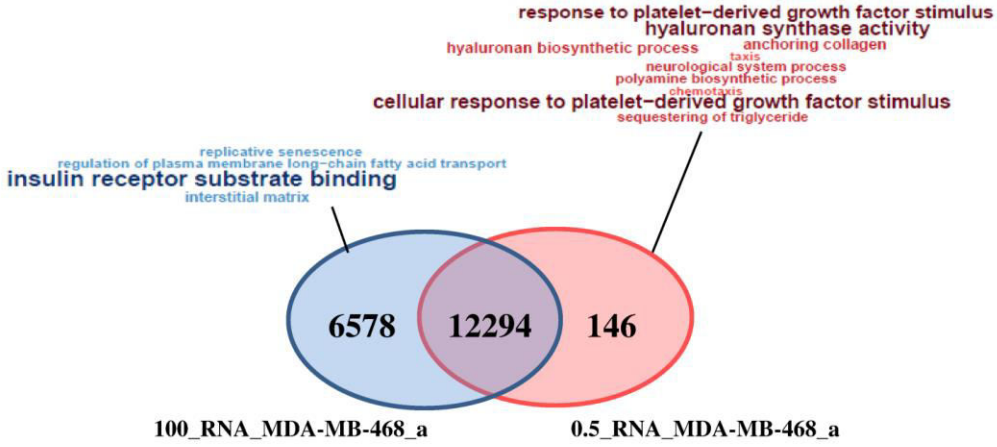
1. Zujewski JA, Kamin L. Trial assessing individualized options for treatment for breast cancer: the TAILORx trial. Future Oncol 2008;4:603–10.
2. Cardoso F, Piccart-Gebhart M, Van't Veer L, Rutgers E. TRANSBIG Consortium. The MINDACT trial: the first prospective clinical validation of a genomic tool. Mol Oncol 2007;1:246–51.
3. Cardoso F, Van't Veer L, Rutgers E, Loi S, Mook S, Piccart-Gebhart MJ. Clinical application of the 70-gene profile: the MINDACT trial. J Clin Oncol 2008;26:729–35.
4. Wong WB, Ramsey SD, Barlow WE, Garrison LP Jr, Veenstra DL. The value of comparative effectiveness research: projected return on investment of the RxPONDER trial (SWOG S1007). Contemp Clin Trials 2012;33:1117–23.
5. Niikura N, Liu J, Hayashi N, Mittendorf EA, Gong Y, Palla SL, et al. Loss of human epidermal growth factor receptor 2 (HER2) expression in metastatic sites of HER2-overexpressing primary breast tumors. J Clin Oncol 2012;30:593–9.
6. Dupont Jensen J, Laenkholm AV, Knoop A, Ewertz M, Bandaru R, Liu W, et al. PIK3CA mutations may be discordant between primary and corresponding metastatic disease in breast cancer. Clin Cancer Res 2011;17:667–77.
7. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med 2012;366:883–92.
8. Bhatia S, Frangioni JV, Hoffman RM, Iafrate AJ, Polyak K. The challenges posed by cancer heterogeneity. Nat Biotechnol 2012;30:604–10.
9. Bidard FC, Peeters DJ, Fehm T, Nolé F, Gisbert-Criado R, Mavroudis D. Clinical validity of circulating tumour cells in patients with metastatic breast cancer: a pooled analysis of individual patient data. Lancet Oncol 2014;15:406–14.
10. Pierga JY, Bidard FC, Mathiot C, Brain E, Delaloge S, Giachetti S, et al. Circulating tumor cell detection predicts early metastatic relapse after neoadjuvant chemotherapy in large operable and locally advanced breast cancer in a phase II randomized trial. Clin Cancer Res 2008;14:7004–10.
11. Lucci A, Hall CS, Lodhi AK, Bhattacharyya A, Anderson AE, Xiao L, et al. Circulating tumour cells in non-metastatic breast cancer: a prospective study. Lancet Oncol 2012;13:688–95.
12. Alix-Panabières C, Pantel K. Circulating tumor cells: liquid biopsy of cancer. Clin Chem 2013;59:110–8.
13. Danila DC, Pantel K, Fleisher M, Scher HI. Circulating tumour cells as biomarkers: progress toward biomarker qualification. Cancer J 2011;17:438–50.
14. Lianidou ES, Markou A. Circulating tumor cells as emerging tumor biomarkers in breast cancer. Clin Chem Lab Med 2011;49:1579–90.
15. Gradilone A, Raimondi C, Nicolazzo C, Petracca A, Gandini O, Vincenzi B, et al. Circulating tumor cells lacking cytokeratin in breast cancer: the importance of being mesenchymal. J Cell Mol Med 2011;15:1066–70.
16. Theodoropoulos PA, Polioudaki H, Agelaki S, Kallergi G, Saridaki Z, Mavroudis D, et al. Circulating tumor cells with a putative stem cell phenotype in peripheral blood of patients with breast cancer. Cancer Lett 2010;288:99–106.
17. Aktas B, Müller V, Tewes M, Zeitz J, Kasimir-Bauer S, Loehberg CR, et al. Comparison of estrogen and progesterone receptor status of circulating tumor cells and the primary tumor in metastatic breast cancer patients. Gynecol Oncol 2011;122:356–60.
18. Barbazán J, Alonso-Alconada L, Muinelo-Romay L, Vieito M, Abalo A, Alonso-Nocelo M, et al. Molecular characterization of circulating tumor cells in human metastatic colorectal cancer. PLoS One 2012;7:e40476.
19. Powell AA, Talasaz AH, Zhang H, Coram MA, Reddy A, Deng G, et al. Single cell profiling of circulating tumor cells: transcriptional heterogeneity and diversity from breast cancer cell lines. PLoS One 2012;7:e33788.
20. Cann GM, Gulzar ZG, Cooper S, Li R, Luo S, Tat M, et al. mRNA-Seq of single prostate cancer circulating tumor cells reveals recapitulation of gene expression and pathways found in prostate cancer. PLoS One 2012;7:e49144.
21. Heitzer E, Auer M, Gasch C, Pichler M, Ulz P, Hoffmann EM, et al. Complex tumor genomes inferred from single circulating tumor cells by array-CGH and next-generation sequencing. Cancer Res 2013;73:2965.
22. Gasch C, Bauernhofer T, Pichler M, Langer-Freitag S, Reeh M, Seifert AM, et al. Heterogeneity of epidermal growth factor receptor status and mutations of KRAS/PIK3CA in circulating tumor cells of patients with colorectal cancer. Clin Chem 2013;59:252.
23. Sieuwerts AM, Mostert B, Bolt-de Vries J, Peeters D, de Jongh FE, Stouthard JM, et al. mRNA and microRNA expression profiles in circulating tumor cells and primary tumors of metastatic breast cancer patients. Clin Cancer Res 2011;17:3600–18.
24. Fehm T, Becker S, Duerr-Stoerzer S, Sotlar K, Mueller V, Wallwiener D, et al. Determination of HER2 status using both serum HER2 levels and circulating tumor cells in patients with recurrent breast cancer whose primary tumor was HER2 negative or of unknown HER2 status. Breast Cancer Res 2007;9:R74.
25. Fehm T, Müller V, Aktas B, Janni W, Schneeweiss A, Stickeler E, et al. HER2 status of circulating tumor cells in patients with metastatic breast cancer: a prospective, multicenter trial. Breast Cancer Res Treat 2010;124:403–12.
26. Ignatiadis M, Xenidis N, Perraki M, Apostolaki S, Politaki E, Kafousi M, et al. Different prognostic value of cytokeratin-19 mRNA positive circulating tumor cells according to estrogen receptor and HER2 status in early-stage breast cancer. J Clin Oncol 2007;25:5194–202.
27. Ignatiadis M, Kallergi G, Ntoulia M, Perraki M, Apostolaki S, Kafousi M, et al. Prognostic value of the molecular detection of circulating tumor cells using a multimarker reverse transcription-PCR assay for cytokeratin 19, mammaglobin A, and HER2 in early breast cancer. Clin Cancer Res 2008;14:2593–600.
28. April C, Klotzle B, Royce T, Wickham-Garcia E, Boyaniwsky T, Izzo J, et al. Whole-genome gene expression profiling of formalin-fixed, paraffin-embedded tissue samples. PLoS One 2009;4:e8162.
29. Du P, Kibbe WA, Lin SM. Lumi: a pipeline for processing Illumina microarray. Bioinformatics 2008;24:1547–8.
30. Andreopoulou E, Yang LY, Rangel KM, Reuben JM, Hsu L, Krishnamurthy S, et al. Comparison of assay methods for detection of circulating tumor cells in metastatic breast cancer: AdnaGen AdnaTest BreastCancer Select/Detect™ versus Veridex CellSearch™ system. Int J Cancer 2012;130:1590–7.
31. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol 2009;27:1160–7.
32. Krebs MG, Metcalf RL, Carter L, Brady G, Blackhall FH, Dive C. Molecular analysis of circulating tumour cells-biology and biomarkers. Nat Rev Clin Oncol 2014;11:129–44.
33. Pantel K, Alix-Panabières C, Riethdorf S. Cancer micrometastases. Nat Rev Clin Oncol 2009;6:339–51.
34. Chui MH. Insights into cancer metastasis from a clinicopathologic perspective: epithelial-mesenchymal transition is not a necessary step. Int J Cancer 2013;132:1487–95.
35. Kalluri R, Weinberg RA. The basics of epithelial-mesenchymal transition. J Clin Invest 2009;119:1420–8.
36. Yu M, Bardia A, Wittner BS, Stott SL, Smas ME, Ting DT, et al. Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. Science 2013;339:580–4.
37. Bidard FC, Weigelt B, Reis-Filho JS. Going with the flow: from circulating tumor cells to DNA. Sci Transl Med 2013;5:207ps14.
38. Chen CL, Mahalingam D, Osmulski P, Jadhav RR, Wang CM, Leach RJ, et al. Single-cell analysis of circulating tumor cells identifies cumulative expression patterns of EMT-related genes in metastatic prostate cancer. Prostate 2013;73:813–26.

Supplemental Data Figure 1

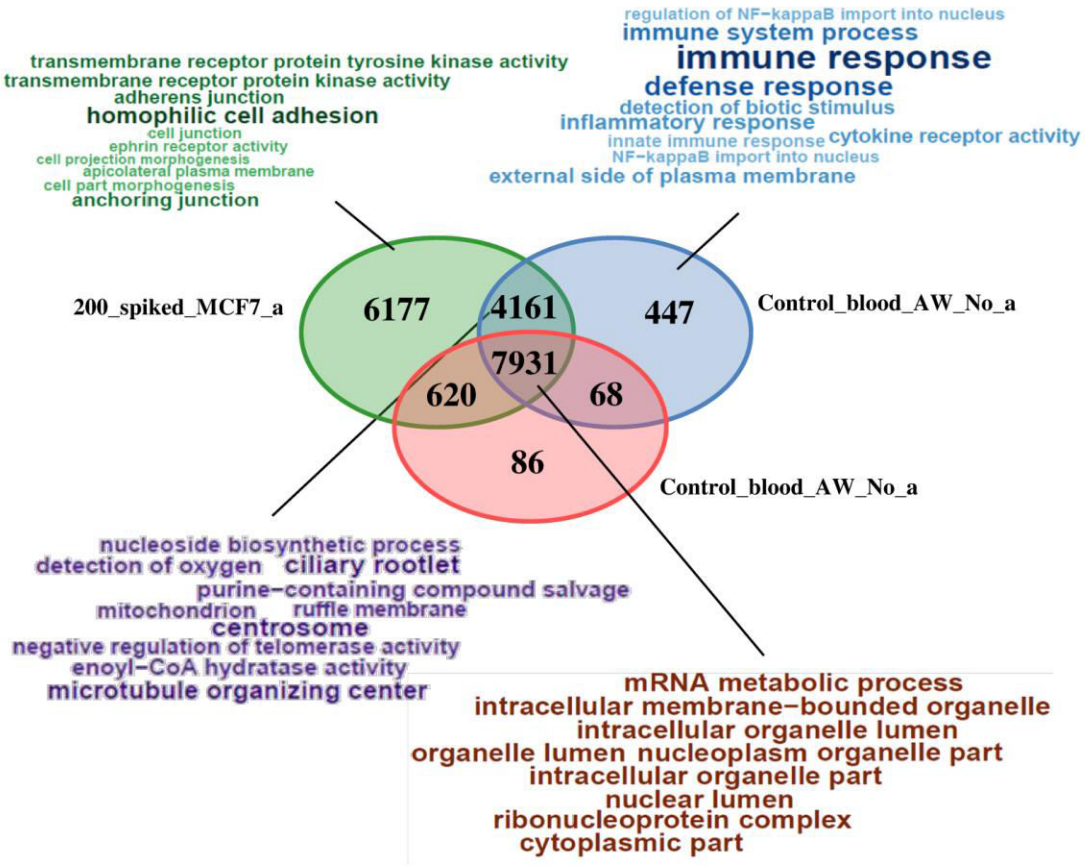
A



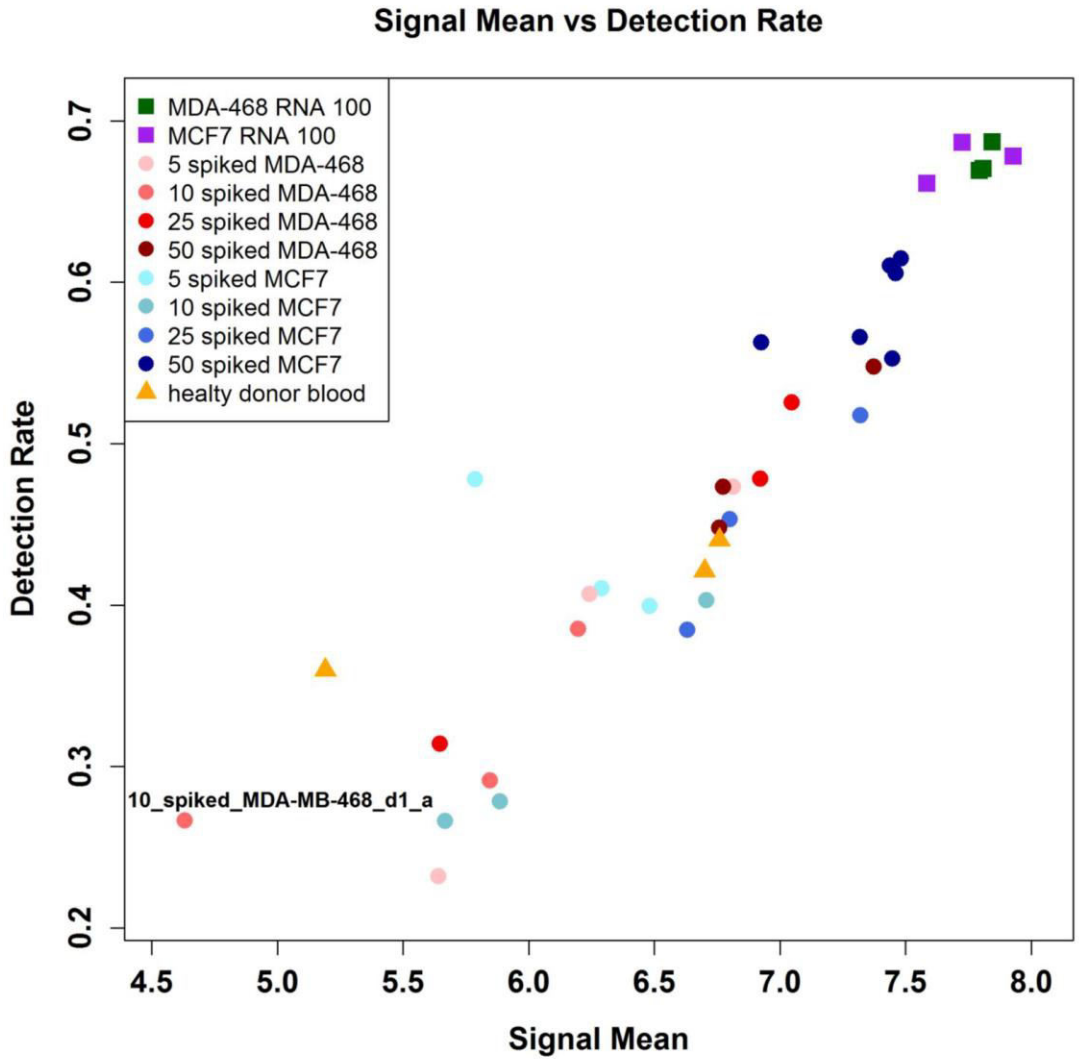
B



Supplemental Data Figure 2



Supplemental Data Figure 3



Supplemental Data Table 1: Patients clinico-pathologic features and CTC status/count data

Patient ID	Approach used for CTC detection	n° CTCs/ 5 mL blood	Primary tumor/ Metastatic lesions			Stage (AJCC)	Line of treatment
			ER status **	PR status **	HER2 status ***		
CTC001	CellSearch	203	Pos/Pos	Neg/Neg	Neg/Neg	T3N1M0	III
CTC002	CellSearch	7	Pos/Pos	Pos/Pos	Neg/Neg	T2N1M0	I
CTC003	AdnaGen	*	Pos/Pos	Pos/Pos	Neg/Neg	T1N0M0	I
CTC004	CellSearch	13	Pos/Pos	Neg/Neg	Neg/Neg	T2N0M0	II
CTC005	CellSearch	0	Pos/Pos	Neg/Neg	Neg/Neg	T1N0M0	I
CTC006	CellSearch	13	Pos/Pos	Pos/Pos	Neg/Neg	T3N0M0	II
CTC007	CellSearch	19	Pos/Pos	Pos/Pos	Neg/Neg	T1N1M0	III

* CTC negative by AdnaTest

** evaluated by immunohistochemistry (ER, cut off= 10% positive cells; PR, cut off_10% positive cells) according to Hammond M.E. et al . J Clin Oncol 2010; 28:2784-95.

*** evaluated by immunohistochemistry and FISH (for HER2 2+ cases)

Supplemental Data Figure 1

Panel A. Pair plots showing correlations between selected samples: Gene expression intensities obtained from samples containing 100 ng of RNA extracted from MDA-MB-468 cells considered as benchmarks and correlated with intensities obtained in profiles using decreasing RNA inputs, respectively 10 ng, 1 ng and 0.5 ng.

Panel B. Eulero-Venn diagrams highlighting the numbers of common and exclusive genes detected in samples derived from 100 ng of MDA-MB-468 RNA and from 0.5 of RNA from the same cell line. Gene Ontologies for exclusive genes in each set are reported using character sizes and color intensities directly proportional the p value for to the gene set enrichment.

Supplemental Data Figure 2

Eulero-Venn diagrams highlighting the numbers of common and exclusive genes detected in samples derived from 200 MCF7 cells spiked into healthy donor blood (HDB); from HDB not treated with AdnaWash buffer and HDB samples treated with AdnaWash buffer. Gene Ontologies for exclusive genes in each set are reported using character sizes and color intensities directly proportional the p value for to the gene set enrichment.

Supplemental Data Figure 3

Distribution plot of signal mean intensities in gene expression data derived from 5-10-25-50 MDA-MB-468 (red dots) or MCF7 (blue dots) cells spiked into healthy donor blood (HDB), from cell-free HDB (yellow triangles), and from 100 ng of RNA isolated from MCF-7 and MDA-MB-468 cells (purple and green squares, respectively) profiled with the Illumina Human Whole-Genome DASL HT.

IJBM
eISSN 1724-6008

Int J Biol Markers 2015; 00(00): e000-e000
DOI: 10.5301/jbm.5000166

SHORT COMMUNICATION

Did circulating tumor cells tell us all they could? The missed circulating tumor cell message in breast cancer

Emanuela Fina[1], Carolina Reduzzi[1], Rosita Motta[1], Serena Di Cosimo[2], Giulia Bianchi[2], Antonia Martinetti[2], Janine Wechsler[3], Vera Cappelletti[1], Maria G. Daidone[1]

[1] Biomarkers Unit, Department of Experimental Oncology and Molecular Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan - Italy
[2] Medical Oncology Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan - Italy
[3] ScreenCell, 75012 Paris - France

Emanuela Fina and Carolina Reduzzi contributed equally to this work

ABSTRACT

Purpose: To compare circulating tumor cell (CTC) detection rates in patients with early (M_0) and metastatic (M_+) breast cancer using 2 positive-selection methods or size-based unbiased enrichment.
Methods: Blood collected at baseline and at different times during treatment from M_0 patients undergoing neo-adjuvant therapy and from M_+ women starting a new line of treatment was processed in parallel using AdnaTest EMT-1/ and EMT-2/Stem CellSelect/Detect kits or ScreenCell Cyto devices. CTC positivity was defined according to the suggested cutoffs and cytological parameters, respectively.
Results: Higher CTC detection rates were obtained with the AdnaTest approach when using for CTC-enrichment antibodies against ERBB2 and EGFR in addition to MUC1 and the classical epithelial surface marker EPCAM (13% vs. 48%). In M_0 patients mainly, CTC positivity rates further increased when EMT- and stemness-related marker expression (PIK3CA, AKT2 and ALDH1) was evaluated in addition to EPCAM, MUC1 and ERBB2. When the physical properties of tumor cells were exploited, CTCs were detected at higher percentages than with positive-selection-based methods, without any difference between clinical stages (78% in M_0 vs. 72% in M_+ cases at baseline). Circulating tumor microemboli (CTMs) were detected in addition to single CTCs with significantly higher frequency in M_0 than M_+ samples (78% vs. 27%, p = 0.0002).
Conclusions: Different approaches for CTC detection probably identify distinct tumor cell subpopulations, but need technical standardization before their clinical validity and biological specificity may be adequately investigated. The distinct role of CTMs compared with CTCs as prognostic and predictive biomarkers represents a further challenge.

Keywords: Breast cancer, Circulating tumor cells (CTCs), Circulating tumor microemboli (CTM)

Introduction

The number of new methods for circulating tumor cell (CTC) detection is rising vertiginously, but few products or devices designed for CTC evaluation have provided robust and independently validated clinical results (1). The CellSearch™ (Janssen Diagnostics, LCC, Raritan, NY, USA) approach was granted FDA approval for CTC detection and monitoring in metastatic solid tumors, but although it produced ample evidence on the prognostic role of CTCs in metastatic (2) and early breast tumors (3-5), CTC enrichment with this approach is limited to epithelial cells and no CTCs are found in a variable rate (35%-55%) of patients with metastatic disease.

In large trials such as SWOG500 (6), a change of treatment based on CTC status did not benefit the patient, which questions the actual biological role of the CTC subpopulation defined according to the CellSearch criteria. Different explanations have been given for such a failure (7), including adequacy of treatment, but maybe it is time to concentrate our efforts on the approx. 50% CTC-negative subset of patients with metastatic breast cancer, which might contain a CTC population undetectable with CellSearch (8). The present work focuses on the issue of going beyond the simple presence/enumeration of CTCs so that the molecular peculiarities of "missed"

Received: May 6, 2015
Accepted: July 13, 2015
Published online: August 26, 2015

Corresponding author:
Vera Cappelletti
Biomarkers Unit, Department of
Experimental Oncology and Molecular Medicine
Fondazione IRCCS Istituto Nazionale dei Tumori
Via Amadeo 42, 20133 Milan, Italy
veracappelletti@istitutotumari.mit.it

© 2015 Wichtig Publishing

CTCs can be identified and further pieces of clinically important information can be unraveled.

Methods

Case series

Patients with histologically confirmed breast cancer and no evidence of metastatic disease (M_0) who were receiving anthracycline/taxane-based neoadjuvant therapy and trastuzumab if HER2 positive, and breast cancer patients with metastatic disease (M_+) who were starting a new line of systemic treatment (mostly endocrine treatment) were prospectively recruited at the Department of Medical Oncology at Fondazione IRCCS Istituto Nazionale dei Tumori, Milan (INT).

Ethical statement

The study was approved by the ethics committee of INT and written consent was obtained from all patients.

Blood sample collection

Samples of peripheral venous whole blood were drawn from all patients using a 26 G needle and collected in K_3EDTA or K_2EDTA BD Vacutainer tubes, if processed with the AdnaTest (AdnaGen, AG, Langenhagen, Germany) or ScreenCell® Cyto (ScreenCell, Paris, France) kits, respectively. The first blood tube was discarded to minimize the risk of contamination with skin epithelial cells. Samples were stored in the dark at 4°C and processed within 1 hour according to the manufacturer's instructions. For each patient, samples were collected at baseline (before starting treatment) and during treatment at established times: 3 and 6 months after initiation of neoadjuvant treatment and around 4 weeks after mastectomy in M_0 women, and at the beginning of a new line of treatment and at 3 months from treatment start or at progression in M_+ women.

AdnaTest approach

CTC positive enrichment

CTC enrichment with the AdnaTest EMT-1/Stem CellSelect (EMT1) and EMT-2/Stem CellSelect (EMT2) kit was performed in parallel on two 5-mL blood samples from the same patient according to the manufacturer's instructions.

CTC detection

For CTC detection, the expression levels of *EPCAM, MUC1, ERBB2, PIK3CA, AKT2, TWIST1 and ALDH1* transcripts were evaluated by semiquantitative multiplex PCR according to the manufacturer's instructions and using the suggested thresholds for positivity (0.15 ng/µL for *EPCAM, MUC1, ERBB2 and ALDH1 and 0.25 ng/µL for PIK3CA, AKT2 and TWIST1*). Samples with *ACTB* (beta-actin) expression <0.70 ng/µL were excluded from the analysis, whereas those passing the quality control criteria where considered as positive for at least 1 of the above-mentioned markers.

ScreenCell cyto approach

CTC enrichment by size

All blood samples were processed within 1 hour after collection using the ScreenCell® Cyto kit (ScreenCell, Paris, France) according to the manufacturer's instructions (9). Briefly, blood was diluted in 4 mL of red blood cell lysis and fixation buffer and incubated 8 minutes at room temperature. Three filtrations of 3 mL of blood each were separately performed for each patient; microporous membranes were rinsed with PBS, collected from the device, air-dried and immediately stained at room temperature for 1 minute with Hematoxylin Solution S (Merck, Darmstadt, Germany) and then for 30 seconds with Shandon Eosin Y Aqueous Solution (Thermo Fisher Scientific Inc., Waltham, MA, USA). Microporous membranes were stored at -20°C until cytological evaluation by a certified pathologist (JW).

Cytomorphological analysis and CTC counts

All membranes were analyzed by the same pathologist (JW) without knowledge of the clinical data. Major criteria for CTC identification were a high nuclear-to-cytoplasmic ratio (≥0.75) and large nuclear size (≥20 µm), whereas minor criteria included irregular nuclear contours and nuclear hyperchromatism. The cytomorphological analysis and CTC count were based on the previously reported criteria of malignancy (10). Circulating tumor microemboli (CTMs) were defined as clusters of at least 2 CTCs, often mixed with platelets and various leukocytes, showing criteria of malignancy like those described for single CTCs. The nuclear-to-cytoplasmic ratios between single CTCs and CTC aggregates are similar (11).

Results were expressed as numbers of CTCs and CTMs for single membranes. For each patient, total CTC or CTM numbers derived from 3 membranes (corresponding to 9 mL of blood) were added together to better meet the criteria for accurate detection of rare events following the Poisson probability distribution (12). Membranes showing poor quality of cytology, estimated on the basis of poor preservation of the leukocytes, were excluded from the analysis. Samples were rated as CTC or CTM positive if at least 1 CTC or CTM was detected in the 3 membranes.

Statistical analysis

Data were analyzed using conventional descriptive statistics. Contingency tables were analyzed by Fisher's exact test and the quantitative measure of the agreement between categorical variables was evaluated by Cohen's kappa statistics, κ. Differences among ordinal variables were assessed by Mann-Whitney's *U* test. Two-tailed $p \leq 0.05$ was adopted as the significance threshold.

Results

The AdnaTest EMT-1/Stem Cell Select + Detect (hereafter referred to as EMT1) and the AdnaTeast EMT-2/Stem Cell Select + Detect (hereafter referred to as EMT2) kits were used

© 2015 Wichtig Publishing

for a head-to-head comparison of CTC identification methods using 2 distinct blood samples from the same patient.

Data on CTC status were separately analyzed for 21 blood samples derived from 13 nonmetastatic breast cancer patients (M_0) undergoing neoadjuvant treatment with conventional anthracycline/taxane schemes, and for 9 samples derived from 6 women with metastatic disease (M_+) who were starting systemic treatment.

An increase in the percentage of samples defined as CTC positive was observed in samples processed using the EMT2 kit, where immunomagnetic enrichment of CTCs takes advantage of the addition of antibodies against EGFR and ERBB2 besides the classical EPCAM/MUC1 antibodies employed in the EMT1 kit (Tab. I). In the overall series, CTC positivity rates roughly passed from 13% to 47%. Such an increase in the number of samples defined as CTC positive was also observed by stratifying patients according to the clinical setting.

Consistent with the above reported observations, there was indeed poor overall agreement between EMT1 and EMT2 on CTC status (κ = 0.159; 95% confidence limits -0.159-0.417). Only 60% of samples were concordantly rated for CTC status. The discordant samples were enriched in EMT1-/EMT2+ cases (11 EMT1-/EMT2+ versus 1 EMT1+/EMT2-).

These results clearly show that, by using the EMT1 kit, 10 samples (6 M_0 and 4 M_+) would have been classified as CTC negative. Such missed CTCs, whose clinical relevance needs to be defined, deserve careful molecular characterization.

To comply with the reported heterogeneity of the CTC population (8, 13), different CTC subpopulations identified by distinct gene expression patterns were separately evaluated. To this end, besides the classical epithelial-like/tumor-associated phenotype for CTC definition (*EPCAM, MUC1, ERBB2*), we also considered cell subpopulations expressing EMT-related genes (*PIK3CA, AKT2, TWIST1*) and stemness-related genes (*ALDH1*).

Overall, considering the optimal CTC selection with EMT2, an increase in positivity rates was observed when adding the expression of EMT-related genes (from 20% to 40%, NS), which reached 47% when expression of *ALDH1* was also included (Tab. II). The identification of CTCs including genes associated with EMT and with stemness, compared with epithelial-like and tumor-associated markers only, increased the positivity rates in the overall population, although the increase was more evident in M_0 than M_+ patients.

In a distinct case series of 14 M_0 and 18 M_+ breast cancers, CTCs and CTMs were identified based on cytological criteria only, after unbiased CTC enrichment by size-selection. The results, expressed as total number of CTCs and CTMs in 9 mL of blood, are reported in Table III using 1 CTC and 1 CTM in 9 mL of blood as positivity cutoff.

CTCs were identified in a high percentage of samples, ranging from 72% to 92% without any differences between

TABLE I - CTC positivity in blood samples processed in parallel with EMT1 and EMT2 kits

Clinical stage	N	Overall CTC detection	
		EMT1+	EMT2+
M_0	21	2	8
M_+	9	2	6

CTC = circulating tumor cells; M_0 = nonmetastatic breast cancer; M_+ = metastatic breast cancer.

TABLE II - EMT2 CTC positivity rates according to different CTC identification criteria

Clinical stage	Criteria for CTC identification		
	Epithelial/tumor-associated genes*	Epithelial/tumor-associated + EMT-related** genes	Epithelial/tumor-associated + EMT-related + stemness-related*** genes
M_0	1/21	6/21	8/21
M_+	5/9	6/9	6/9

CTC = circulating tumor cells; M_0 = nonmetastatic breast cancer; M_+ = metastatic breast cancer.
* *EPCAM, MUC1, ERBB2.*
** *PIK3CA, AKT2, TWIST1.*
*** *ALDH1A.*

TABLE III - CTC and CTM positivity rates in samples processed by size-based selection

Clinical stage	Baseline			During treatment			Any time		
	N	CTC*	CTM*	N	CTC*	CTM*	N	CTC*	CTM*
M_0	14	11 (78)	11 (79)	13	12 (92)	10 (77)	27	23 (85)	21 (78)
M_+	18	13 (72)	5 (28)	12	11 (92)	3 (25)	30	24 (80)	8 (27)
Overall	32	24 (75)	16 (50)	25	23 (92)	13 (52)	57	47 (82)	29 (62)

CTC = circulating tumor cells; CTM = circulating tumor microemboli; N = number; M_0 = nonmetastatic breast cancer; M_+ = metastatic breast cancer.
*Numbers of CTC+ or CTM_+ samples (positivity percentage). Positivity cutoffs ≥1 CTC/9 mL of blood, ≥1 CTM/9 mL of blood.

© 2015 Wichtig Publishing

clinical stages and between samples collected at baseline or during treatment. CTMs, too, were present at similar rates in samples derived from untreated or treated women; however, overall the CTM presence rates were different (p = 0.0002) between M_+ (27%) and M_0 (78%) patients, and the difference was perceptible at baseline (p = 0.0113) and during treatment (p = 0.017).

Median CTC counts did not differ between M_+ and M_0 patients (9 vs. 8.5 at baseline, 10 vs. 15 during treatment, and 9 vs. 11 overall), whereas median CTM counts were higher in patients without clinical evidence of metastases at all tested times (3 vs. 0, p = 0.0056 at baseline; 2 vs. 0, p = 0.0278 during treatment; and 2 vs. 0, p = 0.001 overall). These results suggest early dissemination of the disease and highlight the importance of better knowledge of the molecular heterogeneity of CTCs for predicting progression.

Discussion

CTCs, enriched and identified by their epithelial features, have reached a high level of evidence as prognostic tools in different clinical stages. However, we would like to give further importance to previous observations that, besides the purely epithelial CTCs, there is an additional, missed CTC population (8) whose clinical relevance is still unexplored and deserves further attention. Using immunomagnetic CTC enrichment in 2 small case series in the present study, we were able to demonstrate that by i) optimizing the antibody composition of cocktails used for CTC enrichment by immunobeads and ii) shifting the CTC identification criteria from purely epithelial features (CellSearch criteria) to include also mesenchymal and stem cell features, a significantly higher number of blood samples was defined as CTC positive. Optimized enrichment and less strict CTC definition reduce the differences between clinical stages, suggesting that in patients without clinical evidence of metastases CTCs may show mesenchymal and stemness traits and lose epithelial features.

The extreme approach of unbiased CTC enrichment by a size-based filtration method yields a further increase in CTC positivity rates, paradoxically abolishing the differences in CTC status and numbers between M_0 and M_+ patients. This raises interesting questions about distinct molecular features of CTCs in the 2 settings, which might not only depend on the stage of the disease but also help in predicting progression. Whereas this is a possible explanation, the absence of standardization in the CTC assay (with the exception of the CellSearch approach) delays the achievement of clinical validation and represents a limit of the present study.

CTMs, representing heterogeneous clusters of CTCs with blood cells and platelets, are an additional, promising feature worth being explored. It was recently suggested in animal models that CTC clusters are more proficient in generating metastases than single CTCs (14): from such a perspective, the significantly lower number of CTMs in M_+ compared to M_0 patients represents a new finding that (although apparently not intuitive, similarly to the increases in CTC positivity reported above for improved detection methods) supports the necessity to uncover the message of these missed CTCs. We also speculate that CTMs are particularly worth characterizing from the molecular point of view in M_0 patients, whereas in M_+ patients single CTCs that could derive from metastatic sites may harbor the molecular makeup of the metastasis, recapitulating the clonal evolution of the disease.

It is important to mention that the currently employed technical approaches for CTC detection, albeit based on distinct properties, share a lack of biological specificity. Whereas it can be affirmed that the identified CTCs are by all criteria cancer cells, data on their ability to invade, proliferate or cause metastases are as yet scanty. A different way of investigating the clinical role of such cells is an urgent need if we want to fulfill the promises of the liquid biopsy approach.

Acknowledgment

The authors would like to thank all patients who participated in the study and their families.

Disclosures

Financial support: Financial support was received from Associazione Italiana per la Ricerca sul Cancro, AIRC (IG 10611 - M.G. Daidone); European Commission under the 7th Framework Programme, grant agreement No. 260791 Eurocan Platform; Italian Ministry of Health. E.F. is the recipient of an AIRC fellowship.
Conflict of interest: Janine Wechsler is a consultant pathologist at ScreenCell. Maria Grazia Daidone is a member of the *International Journal of Biomarkers* Editorial Board. The authors confirm that her involvement as editor of the journal in no way alters the adherence to the journal policies.

References

1. Alix-Panabières C, Pantel K. Challenges in circulating tumour cell research. Nat Rev Cancer. 2014;14(9):623-631.
2. Cristofanilli M, Budd GT, Ellis MJ, et al. Circulating tumor cells, disease progression, and survival in metastatic breast cancer. N Engl J Med. 2004;351(8):781-791.
3. Pierga JY, Bidard FC, Mathiot C, et al. Circulating tumor cell detection predicts early metastatic relapse after neoadjuvant chemotherapy in large operable and locally advanced breast cancer in a phase II randomized trial. Clin Cancer Res. 2008; 14(21):7004-7010.
4. Lucci A, Hall CS, Lodhi AK, et al. Circulating tumour cells in non-metastatic breast cancer: a prospective study. Lancet Oncol. 2012;13(7):688-695.
5. Rack B, Schindlbeck C, Jückstock J, et al; SUCCESS Study Group. Circulating tumor cells predict survival in early average-to-high risk breast cancer patients. J Natl Cancer Inst. 2014;106(5):dju066.
6. Smerage JB, Barlow WE, Hortobagyi GN, et al. Circulating tumor cells and response to chemotherapy in metastatic breast cancer: SWOG S0500. J Clin Oncol. 2014;32(31):3483-3489.
7. Raimondi C, Gradilone A, Naso G, Cortesi E, Gazzaniga P. Clinical utility of circulating tumor cell counting through CellSearch(®): the dilemma of a concept suspended in Limbo. Onco Targets Ther. 2014;7:619-625.
8. Yu M, Bardia A, Wittner BS, et al. Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. Science. 2013;339(6119):580-584.

© 2015 Wichtig Publishing

9. Desitter I, Guerrouahen BS, Benali-Furet N, et al. A new device for rapid isolation by size and characterization of rare circulating tumor cells. Anticancer Res. 20011;31(2):427-441.
10. Hofman VJ, Ilie MI, Bonnetaud C, et al. Cytopathologic detection of circulating tumor cells using the isolation by size of epithelial tumor cell method: promises and pitfalls. Am J Clin Pathol. 2011;135(1):146-156.
11. Cho EH, Wendel M, Luttgen M, et al. Characterization of circulating tumor cell aggregates identified in patients with epithelial tumors. Phys Biol. 2012;9(1):016001.
12. Allan AL, Keeney M. Circulating tumor cell analysis: technical and statistical considerations for application to the clinic. J Oncol. 2010;2010:426218.
13. Barriere G, Fici P, Gallerani G, Fabbri F, Zoli W, Rigaud M. Circulating tumor cells and epithelial, mesenchymal and stemness markers: characterization of cell subpopulations. Ann Transl Med. 2014;2(11):109.
14. Aceto N, Bardia A, Miyamoto DT, et al. Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. Cell. 2014;158(5):1110-1122.

© 2015 Wichtig Publishing